NO. 1
Price 25¢

SOME POSSIBLE CONTRIBUTIONS OF STATISTICAL QUALITY CONTROL TO ENGINEERING ECONOMY

by

Eugene L. Grant
Professor of Economics of Engineering
Stanford University
Stanford, California

for presentation at the

FOURTH NATIONAL CONVENTION
and
FIFTH MIDWEST CONFERENCE

of the

AMERICAN SOCIETY for Quality Control

June 1 and 2, 1950
Milwaukee Auditorium, Milwaukee, Wisconsin

FIFTH MIDWEST QUALITY CONTROL CONFERENCE
P. O. Box 1204, Milwaukee 1, Wisconsin

SOME POSSIBLE CONTRIBUTIONS OF STATISTICAL QUALITY CONTROL TO ENGINEERING ECONOMY

Presented before American Society for Quality Control at Milwaukee, Wisconsin, June 1, 1950 by Eugene L. Grant, Professor of Economics of Engineering, Stanford University, Fellow ASQC, Member San Francisco Bay Area Section ASQC

A considerable fraction of the numerous decisions required in industry are made as intuitive judgments; they are "hunch" decisions. Such decisions are often made without any conscious attempt to define the alternatives that are available. Even though alternatives are defined, it is common for no effort to be made to express the prospective differences between the alternatives in commensurable units.

In contrast, some decisions in industry are based on engineering economy studies. In an economy study, an attempt is made to discover all of the promising alternatives and to arrive at a clear definition of the alternatives to be compared. Estimates are made of the prospective consequences of choosing each of the alternatives. In so far as possible, these expected consequences are made commensurable by being expressed in terms of prospective money disbursements and, whenever appropriate, prospective money receipts. Where proposed investments are involved, the prospective money time series are made comparable by calculations that reflect the time value of money. Finally a choice among the alternatives is made that gives weight not only to their money differences but also to those prospective differences that have not been reduced to money terms.

Many decisions in industry relate in some way to matters connected with quality. It is my impression that the techniques of engineering economy have been used less frequently in quality matters than in a number of other arease of decision making. No doubt one reason for this failure to make economy studies is the fact that decisions on quality matters are so numerous that it is felt that little time can be devoted to each decision. Another reason is the difficulty of placing a money valuation on many of the differences between alternatives that require consideration in dealing with quality matters.

How can statistical quality control contribute to better decisions on matters related to the economic aspects of quality? This contribution may be made in two ways. One of these ways is through the provision of better alternatives than were previously available in dealing with certain problems involving quality. Familiar examples of such better alternatives are

the use of the Shewhart control chart for controlling quality during production and the use of acceptance procedures such as are described in the Dodge-Romig tables and the Navy tables. The other way in which statistical quality control contributes to better decisions is through the provision of better methods for estimating the consequences of the various alternatives.

In the initial stages of the application of statistical quality control methods in any organization, many changes in existing procedures will be proposed. The advantages of a proposed change in procedure will often be so great that the economic merits of the proposal will be clear without any formal evaluation in an economy study. In the following discussion, Cases A and B are intended to represent this familiar situation. Later on, after the more obvious opportunities for economies through statistical quality control have been exhausted, further applications may require careful economy studies; this is illustrated in Cases C and D.

<u>Case A. Routine Use of the Control Chart Provides an Improved Alternative.</u> A certain measurable quality characteristic of a manufactured product is influenced by one or more adjustments of a machine used in the manufacturing process. From time to time one article just off the machine is measured by the operator or by an inspector. If the measured value of the quality characteristic of this one article is either above or below the aimed-at value, the machine is adjusted accordingly.

This case describes a condition that often exists where manufacturing is conducted without benefit of statistical quality control. For example, the quality characteristic in question might be the weight of the contents of a food container, the shear strength of a spotweld, the inside perimeter of a paper bag, or the powder charge in a round of ammunition. Anyone familiar with the techniques of statistical quality control recognizes that the frequent machine adjustments made by the operator will inevitably increase the variability of the process. The routine use of the control chart provides an improved method of determining when machine adjustments shall be made; the use of $\bar{X}$ and R charts can be coupled with instructions that the process is to be left alone unless a point falls outside of control limits. The monetary advantages of the resulting quality improvements are sometimes spectacular.

Case A is representative of the frequent circumstance in which the improved alternative to existing practice is provided by the routine systematic use of one of the tools of statistical quality control.

Case B. Statistical Analysis Makes Possible the Evaluation of the Consequences of Different Alternatives. A certain quality characteristic of a manufactured article is examined by a destructive test that cannot be completed until several days after the production of the article. A substantial number of articles are tested from each day's production. The exact order of production is preserved until the results of the tests are known. Whenever an article tested fails to meet the quality specification, all of the articles are scrapped that were produced between the last preceding test specimen that met the quality specification and the next succeeding test specimen that met the specification. This acceptance procedure results in the scrapping of more than five per cent of the product. It also involves substantial testing costs, including the costs of the items destroyed by the testing procedure itself.

The introduction of the techniques of statistical quality control into this manufacturing organization is accompanied by a statistical analysis of a large volume of past test results. This analysis, using control charts and other statistical tests, indicates that the test data are consistent with the hypothesis that a constant system of chance causes has been operating during most of the period of record. In other words, most of the articles scrapped did not differ materially from the articles accepted. The articles accepted have, generally speaking, proved satisfactory for the purpose intended. This analysis indicates that there should be a thorough review of the production methods, the quality specification, and the acceptance procedure.

Meanwhile a decision must be made regarding the acceptance procedure to be used in the immediate future. One suggestion is to continue the past program of testing until this thorough review of the matter is completed but to make an immediate change in the acceptance criteria. It is suggested that all of the product, rather than only part of it, should be accepted unless there is some specified change in the average test value or in the variability of the test results. Not much of an economy study is needed to show that this suggestion is superior to continuing the present acceptance criteria. As the statistical analysis of the past results has indicated that the rejected product was as good as the accepted product, and as the accepted product has generally been satisfactory, it is clearly more economical to accept all the product than to continue rejections on the present basis.

This immediate decision in Case B is representative of the frequent circumstance in which the consequences of two

proposed alternatives affecting quality cannot be evaluated without statistical analysis.

Case C. Are Extra Production Costs Justified by the Prospect of Fewer Lot Rejections? This case might be described as the case of the statistically sophisticated vendor versus the statistically unsophisticated purchaser. A certain quality characteristic of a manufactured article is tested by a destructive test. The purchaser's specification states that the minimum acceptable value of this characteristic is 1,200 units. (The characteristic in question may be thought of as strength, or as any other characteristic measurable only by a destructive test.) Actual acceptance inspection is on a lot-by-lot basis with acceptance or rejection of a lot dependent only on the evidence of the sample from the particular lot. A single article chosen at random from each lot is tested. If the test value is 1,200 or more, the lot is accepted. If not, a second article is tested. If the second test value also falls below 1,200, the lot is rejected; otherwise it is accepted.

The manufacturer of this product introduces statistical quality control techniques into his plant. A control chart is applied to this particular quality characteristic. With the routine use of the chart, it turns out to be fairly easy to bring this process into a state of statistical control. However, with present production methods the average value is about 1,375 and the standard deviation is about 100. If it is assumed that the distribution is roughly normal, this means that about four per cent of the statistically controlled product will fall below the specified minimum value of 1,200.

This matter is reviewed by the production engineers of the manufacturing company to determine what changes are necessary for all of the product to meet the 1,200 specification. Two alternative measures for improvement are considered. Alternative I is to use a more costly raw material; it is believed that this will raise the average to at least 1,500 with no change in the standard deviation. Alternative II is to purchase certain new equipment; it is believed that this will cut the process dispersion in half so that the standard deviation will be 50 instead of 100. The extra cost of the more expensive material in Alternative I will increase production costs by seven per cent. The investment charges on the new machine in Alternative II will increase production costs by nine per cent.

A review of the costs of these two alternatives leads to a suggestion of Alternative III. This is to make no change in the production process and to continue to submit product

that is, on the average, four per cent defective. It is evident that if this alternative is selected there is only one chance in 25 that a single test specimen will fall below the specification limit of 1,200. With the acceptance procedure in force under which two speciments from a lot must both fall below 1,200 for the lot to be rejected, there is only one chance in 625 that any submitted lot will be rejected. This appears to involve a negligible cost compared to the cost of improving the product quality by either Alternative I or Alternative II. Hence Alternative III is clearly the most economical of the three.

Does Case C Refer to a Common Situation? The economy study in Case C resulted in a deliberate decision to continue to produce and to submit for acceptance a statistically controlled product with four per cent of the product below the specification limit. This decision was based, in part, on an evaluation of the probable frequency of rejections under a weak acceptance procedure. I have personally heard of only one case where a statistical analysis of a production process and of the acceptance procedure resulted in the decision that it was economical to take advantage of the weakness of the acceptance procedure. With the spread of statistical quality control, however, it is inevitable that a great many producers will be faced by this type of decision. Even with the best of sampling acceptance procedures, it is appropriate for a vendor to give consideration to the operating characteristic curve of the sampling acceptance plan and to weigh the risks of lot rejection. Weak sampling acceptance procedures provide a strong economic incentive to submit product with a moderate percentage defective provided it is possible to maintain statistical control of the production process and provided the cost of reducing the percentage of defectives is high in relation to the probable costs associated with rejected lots.

Under What Circumstances Can Satisfactory Results Be Obtained With Weak Acceptance Procedures? The weakness of the acceptance procedure described in Case C is pointed out in nearly all introductory presentations of the basic concepts of statistical quality control. For instance, it is evident that if many lots 50% defective are submitted for acceptance, three-fourths of them will be accepted. Nevertheless, there has been widespread use of this acceptance procedure as well as many others that are somewhat better but are similar in giving little protection against the acceptance of lots containing a substantial percentage of product that does not conform to specifications.

In spite of the limitations of such acceptance schemes when subjected to statistical analysis, we are doubtless all familiar with circumstances where they are used with satis-

factory results. That is, the results are satisfactory in the sense that there is no great complaint about the quality of the articles accepted. It is helpful to consider the different possible circumstances under which such satisfactory results may be obtained. They are as follows:

Either: (1) All of the articles in the submitted lots conform to the quality specification. Here the only effect of the acceptance procedure is to exert pressure on the manufacturer to continue this desirable condition. As implied by Case C, this pressure is strongest if the manufacturer is relatively unsophisticated as far as statistical knowledge is concerned and is therefore unaware of the weakness of the acceptance scheme.

Or: (2) The manufacturing process is badly out of statistical control so that submitted lots are either relatively very good or relatively very bad in their conformity to the quality specification.

Or: (3) The quality specification contains a substantial margin of safety. The lots accepted contain varying proportions of non-conforming articles. However, because of this margin of safety, most of these non-conforming articles, although technically defective according to the specifications, are good enough for the use to which they are put. Where an acceptance procedure is weak, the weakness may be offset, at least in part, by a design specification that requires articles that are better than actually necessary for the purpose at hand.

Often the satisfactory results obtained at present with weak acceptance procedures are a result of (2) lack of statistical control in the product submitted for acceptance, in combination with (3) a large margin of safety in the design specification. As more and more industrial products subject to sampling acceptance inspection are produced under conditions of statistical control, it will become less and less possible to rely on the lack of statistical quality control of submitted product to provide satisfactory results from weak acceptance procedures. Hence improved control of production processes will require improved acceptance procedures. This change will also tend to bring under review a matter that has not yet been given much consideration in many industries, the matter of the economic relationship between design and acceptance specifications.

<u>The Relationship Between Design and Acceptance Specifications.</u> Shewhart, in his early writings, was doubtless the first to emphasize the distinction between a design

specification and an acceptance specification. The design specification, referred to earlier in this paper as the quality specification, involves a statement of what is desired in the various quality characteristics of a product. The acceptance specification deals with the means of judging whether the desired qualities are actually obtained. These two specifications should properly be viewed as interrelated matters.

The greater the margin of safety above what is really needed that the designer puts into his specifications of quality characteristics measurable as variables, the less severe the acceptance procedures need to be in order to secure product that is actually satisfactory for the purpose intended. This statement is not restricted to those circumstances in which weak acceptance procedures are used but is also applicable when the sampling acceptance procedures in use have been developed with the benefit of the statistical approach. In modern acceptance sampling, the starting point for the decision on the acceptance specification is often the determination of a percentage defective to be used as an acceptable quality level or as an average outgoing quality limit. One of the elements in deciding on an AQL or an AOQL is the margin of safety written into the design specification.

Three General Classes of Costs Entering into Economy Studies Involving Quality Matters. In determining design and acceptance specifications, or in making any other decision relative to quality matters, it is advantageous to think of three general classes of costs that may be influenced by quality decisions. These may be somewhat loosely referred to as (1) production costs, (2) acceptance costs, and (3) unsatisfactory-product costs.

In this usage, the expression production costs is intended to refer to those costs involved in the production of the article to which the design specifications and the acceptance specifications apply. Different design specifications may require different materials, different labor skills, different amounts of labor time, and different machines. For example, increased strength requirements for a part may change the material to be used; closer required tolerances on dimensions may call for the use of newer or different machines. This general class of costs properly includes the production expenses on all product that is discarded as not meeting specifications. It also includes sorting costs, if any, on rejected lots, and costs of rework necessary to make product acceptable.

The acceptance costs include not only direct testing and inspection costs but also the costs of administering the acceptance program.

The expression _unsatisfactory-product costs_ is intended to refer to those costs resulting from the acceptance of product that turns out to be unsatisfactory for the purpose intended. It should be recognized that product that is technically defective in the sense of failing to meet design specifications does not necessarily lead to unsatisfactory-product costs if there is a sufficient margin of safety in the design specifications. Of these three classes of costs affected by quality decisions, this class is inherently the most difficult to evaluate.

Doubtless the greatest difficulty in evaluating unsatisfactory-product costs occurs in the consumers' goods industries where the product goes to a great many different customers who do not apply any formal acceptance tests to the product that they purchase. It is hard to predict the consequences to the manufacturer of consumers' goods when some stated percentage of his product fails to give satisfactory service to its purchasers, and it is even more difficult to place a money value on these consequences. Where the consumers' product carries a guarantee, past customer service costs can be used as a guide to judgment, and changes in these costs can be carefully watched and related to changes in design specifications and in inspection and acceptance procedures.

The most favorable circumstances for arriving at a judgment on unsatisfactory-product costs exist where all of the product goes to a single user who has prepared the design and acceptance specifications.

Case D. An Economy Study to Determine Design and Acceptance Specifications. A manufacturer desires to make an economy study to establish design and acceptance specifications for a particular manufactured part where strength is an important quality characteristic. This strength must be determined by a destructive test. As this product is made in batches with the possibility of considerable variation from batch to batch, acceptance or rejection applies to entire batches. The objective of the design and acceptance specifications is that each part shall have a strength of at least 2,000 units, and that this result shall be achieved with a minimum total cost, all things considered.

A number of different possible combinations of design and acceptance specifications are proposed. For purposes of discussion, let us represent these various proposals by two extreme alternative suggestions. Alternative I involves a margin of safety of 1,000 units in the design specification, so that a minimum strength of 3,000 units is required. The

associated acceptance specification is simple. One article is tested from each batch; if the strength is 3,000 or more the batch is accepted; if less, the batch is scrapped.

The proposed acceptance specification in Alternative II involves a control chart for average and range and treats the test results as variables rather than as attributes, giving consideration to dispersion as well as to average value. It makes use of the modern ideas of varying the severity of the acceptance criteria and the amount of the testing with the quality history. Because of the additional protection given by better acceptance procedures, the margin of safety written into the design specification is reduced from 1,000 to 250.

An economy study to compare Alternatives I and II requires consideration of the effect of each on production costs, on acceptance costs, and on unsatisfactory-product costs.

If less costly production methods are required because of the smaller margin of safety in the design specification in Alternative II, this cost saving for Alternative II should be evaluated. The cost of product scrapped because of rejection of batches will need to be estimated for each alternative. This estimate must start with consideration of the expected distribution of the strength values with each quality objective (i.e., with the 3,000 minimum of Alternative I and the 2,250 minimum of Alternative II) and an estimate of the extent to which the process will fail to show statistical control. It is also necessary to take account of the operating characteristic of each acceptance plan.

In considering the acceptance costs influenced by the choice between the alternatives, the total amount of testing under each plan will need to be computed. This can be determined exactly for Alternative I where the amount of testing is fixed as one article per batch; the prediction for Alternative II requires an estimate of the proportion of the time the process will qualify for reduced inspection. It is also necessary to recognize the greater administrative costs associated with the more complicated acceptance scheme of Alternative II.

The evaluation of the unsatisfactory-product costs must start with an estimate of the percentage of unsatisfactory product that will be accepted under each alternative. In this estimate, the proportion of accepted product falling below the design specification limits (3,000 and 2,250 respectively) is of no importance; it is necessary to predict the proportion of product falling below the actual requirement of 2,000 that will be submitted and accepted under each alternative. It is

also necessary to place a money valuation on the costs associated with the acceptance of a unit of unsatisfactory product.

The next step in the economy study is the summation of the three classes of costs for each alternative -- not only for Alternatives I and II but also for any other combinations of design and acceptance specifications that are under consideration. A comparison of the cost sums may eliminate some of the alternatives but may indicate that other alternatives deserve further study before a final decision. As these computed costs depend in part on a particular assumption regarding the centering, dispersion, and stability of the production process, it may be necessary to consider how the costs would behave with several different assumptions about the process.

The final step in the economy study is a choice among the alternatives, giving consideration both to the monetary comparison and to any other relevant differences between the expected consequences of the alternatives that were not expressed in money terms.

<u>Some General Comments on Economy Studies Relating to Design and Acceptance Specifications.</u> Cases A and B described situations familiar to most of you; if parallel conditions have not existed in your own plant, you have known of them in some other plant. The general conditions that might give rise to a Case C are also familiar, even though you may not have knowledge of an economy study that reached the stated conclusion.

In contrast to this, I know of no one who has actually made an economy study such as that described in Case D. This case is intended to suggest a logical development from the increased use of statistical quality control in industry. You may therefore quite properly view Case D with a critical eye.

Several objections might be made to the suggestion that design and acceptance specifications should be viewed as related parts of a single problem of economy and that in many cases economy studies should be made to establish these specifications.

One possible objection is that this requires the designer to get together with the adminstrator of acceptance procedures; when they do get together it is clear that both must approach the matter from the statistical point of view. Although the people who administer acceptance procedures are rapidly acquiring this viewpoint if they do not have it already, this is not necessarily true of designers. In those places where design

engineers continue in a naive and unrealistic view of statistical matters, it will doubtless be impossible to secure satisfactory consideration of design and acceptance specifications as related parts of a single problem.

A second possible objection is that the consequences of various alternative proposals for design and acceptance specifications are difficult to forecast and to express in money terms. This follows the pattern of a common objection to the making of economy studies as a guide to decisions in industry. The answer to this objection is that forecasts of consequences of decisions are the stuff of which decisions should be made. There is an implied forecast in every "hunch" decision, just as there is a conscious forecast in an economy study.

A third objection is that economy studies such as the one suggested in Case D may be complicated and therefore costly. This objection raises an important point; any sorts of studies leading to decisions should pay their own way in the sense that the saving from the resulting improvement in decisions should exceed the cost of the studies. It seems likely, however, that industry affords many opportunities for such savings through improvement in decisions on quality matters.

<u>The Place of the Economy Study in Dealing with Quality Problems.</u> This paper has not attempted to outline all of the ways in which the viewpoint and techniques of statistical quality control can contribute to economy studies involving quality. It has rather attempted to present certain provocative ideas through the medium of a discussion of a few general types of cases. A general conclusion is that, as time goes on, more and more quality problems in industry are likely to be approached by the techniques of engineering economy. The people who undertake such economy studies are likely to have an interesting time as well as contributing to the worthy objective of better industrial product at lower cost.

1

NO. 2
Price 25¢

SPECIFICATION, MEASUREMENT, AND CONTROL OF QUALITY

by

Arthur G. Dalton
Supt. of Quality Control
Western Electric Co.
New York, New York

for presentation at the

FOURTH NATIONAL CONVENTION
and
FIFTH MIDWEST CONFERENCE

of the

AMERICAN SOCIETY for
Quality Control

June 1 and 2, 1950
Milwaukee Auditorium, Milwaukee, Wisconsin

FIFTH MIDWEST QUALITY CONTROL CONFERENCE
P. O. Box 1204, Milwaukee 1, Wisconsin

SPECIFICATION, MEASUREMENT and CONTROL of QUALITY

by

A. G. DALTON Fellow, A.S.Q.C.

Superintendent, Quality Control
WESTERN ELECTRIC COMPANY.

The continuing success of any manufacturer or supplier depends upon how well he keeps his customers satisfied. This means that manufacturers and suppliers must know what will keep their customers satisfied. For some products, high quality is paramount, and in others, low cost is the predominant desire. However, in the vast majority of instances, both quality and cost are important factors in the customer's mind, and both must be controlled at satisfactory levels.

In order to control anything, there must be a clean cut objective or specification, a means of measuring how well the objective is being met, and there must be positive action taken when the objective is not met. Everybody knows what "cost" is - and how to specify it, how to measure it, and how to control it. However, when it comes to specifying, measuring, and controlling quality, we are up against a much more intangible subject.

QUALITY CHARACTERISTICS

Let us consider this word "quality" for a minute. There are two basic and important concepts of quality - one of these applies to a reality entirely independent of relationship to anything else or to our needs and wants. The other, and most common concept of quality, has to do with how we react as a result of contact with this real quality. The latter concept of quality is measurable in terms of the physical properties of things that make them desirable or useful to us. For example - weight, length, speed of operation, etc. It is

necessary that we recognize the difference between these two concepts and concentrate on the latter if we are to satisfy the people who will use our products.

How are we best satisfied? We might say that our opinion of satisfactoriness depends on what we get compared with what we expect to get for the price paid. We look for general satisfactoriness, of course, in a broad sense, but the fact is that adequacy, dependability and cost also tend to influence our overall opinion of the satisfactoriness of things. Naturally the relative importance of these factors will vary quite widely, depending upon the nature of the product; and our engineers, in writing specifications, must have in mind the relative emphasis that should be placed on adequacy, dependability and cost.

In many instances, a high degree of uniformity and interchangeability is of major importance. Even in those cases where the need for interchangeability is not obvious, we still find that we are creatures of habit and react favorably to consistency. We prefer that the things with which we come in contact and with which we have been satisfied in the past, look the same, feel the same, or behave the same each time we come in contact with them. If they differ, to a recognizabledegree, from kindred products with which we have been satisfied in the past, we are inclined to be skeptical of their quality.

Unfortunately, it is impossible to make two or more things exactly alike, however hard we try. When we are striving for customer satisfaction and approval, even in things which might

vary one from another quite a bit without much danger of these differences being detected, we try to keep them as much alike as practicable. When we are dealing with precise measurements and characteristics, and when a high degree of interchangeability is needed, we must call on all available resources to keep the variations to the required minimum. Apart from the customer's needs, too much variability between parts - particularly in complex products - can cause many headaches and unnecessary expense in subsequent assembly operations.

SPECIFICATION

Recognizing the impossibility of making two or more parts exactly alike, design engineers provide tolerances for permissible variations. These margins may be set either to facilitate a subsequent mating of parts in assembly, or to provide some desired quality objective in the finished product. It is essential in this connection, of course, that the establishment of such tolerances include consideration of the capabilities of the tools and facilities which will be used for production. However, requirements of this nature, covering tolerances for individual parts, are not sufficient in themselves. A complete specification of quality objectives should include requirements for the aggregate quality of a multitude of such products, considered as a whole. Such requirements which are often considered as inspection requirements, as contrasted with design requirements, can be specified in a number of ways - the most common being in terms of permissible percent defective - although there are many instances where they can best be specified in terms of the average and distribution.

It is important in this connection that engineers realize that quantitative requirements which cannot be verified, cannot be relied upon to control quality. For example, if minimum point thickness is desired for a protective coating, the design specification must also indicate the conditions of test under which this requirement may be considered to have been met. Supplementary information of this kind is also necessary when requirements cannot be verified because the specified limits for a characteristic represent destruction - as, for example, in the case of a requirement that a fuse must operate under a load of three amperes. Proper specification of requirements of this nature necessitate a knowledge of some of the basic principles upon which quality control techniques are based.

There is one further point that should be stressed in connection with specifications and that is the impracticability of making them entirely self-sufficient. A specification is one of the means by which a Design Engineer attempts to convey the essential elements of the design intent to the people who are going to make and inspect the product. It may even be looked upon as a contract, but we know that a contract is not legally binding unless there is a meeting of the minds of the contracting parties. In order to obtain the most satisfactory results, the man who writes the specification and the man who makes the product must each know what is expected of him. They must both speak the same language. There should be a mutual understanding of the things that will be done during manufacture, even though they may not be included in the specification. For example:

the precautions that will be taken to assure good workmanship, good housekeeping and freedom from dirt, dust, solder splashes, etc., where such things could adversely affect the quality of the product.

There should also be a mutual understanding of what is meant when tolerances for parts are included in the specification. For example: a specification for a shaft might include the following requirements:

Length - 4.000 ± .005".

Diameter - 0.500 ± .001".

Does this mean that a single shaft can be considered as meeting the specification requirements if it is 4.005" long but tapers a little bit so that it measures .501" in diameter at one end and .499" in diameter at the other end? Does it mean that the shaft need not be round, and that two measurements of the diameter taken 90 degrees apart may vary as much as .002"? Or does it mean that all shafts should be round, not oval and that the diameter of each shaft should be practically uniform throughout its length? Does it mean that in any lot of shafts it is all right to have some that are .499" in diameter and others .501" in diameter? The Design Engineer may have one idea and the Manufacturer may have another. This is just an illustration of the need for a common understanding among design, production and inspection personnel and the obvious economic advantages of being sure that this understanding exists before any large amount of product is manufactured.

MEASUREMENT

Let us now go to the second item - the measurement of quality. When it comes to the measurement of quality, we think in terms of inspection and inspection results.

In thinking of inspection we must also think about inspection costs and inspection ratios. However, before we can begin to make intelligent comparisons of inspection costs and inspection ratios among different production units, we must know pretty well what part of the job is classified as "inspection" and what part is classified as "operating". It is quite likely that in certain Shops work, that in other places is regarded as "inspection", is classified as "operating" and the reverse condition may also exist. Furthermore, if, following comparison of inspection ratios, management pressure is exerted to reduce these ratios, there may be a tendency to reclassify operations formerly considered as "inspection", as "operating", with an indicated improvement in inspection ratio. This may tend to relieve pressure where it is needed and it may exert pressure where it is not needed.

The idealistic view would be to consider all work necessary to make product of satisfactory quality as "operating" and to consider "inspection" as that work necessary to verify the fact that the quality of product is satisfactory. Reduced to its simplest terms this would mean that if a part, sub-assembly or product was not of good enough quality to sample, then some part of the work which should be classified as "operating" was not adequately performed.

Let us take a simple example such as soldering wires to terminals. Our ultimate objective, of course, is to get satisfactory soldered connections. We can set up the job in either one of two ways:

(a) We can instruct the operator who actually solders the connections to look at each connection immediately after it has been soldered to see that it is satisfactory before proceeding to solder the next connection. If it is not satisfactory the operator will, of course, re-solder it and make it satisfactory, or

(b) We can have one operator do nothing but soldering and then have another operator look at each connection to see that it is satisfactory and arrange for the correction of any which are found to be unsatisfactory.

Whichever way we do the job, up to this point, we have done just one thing and that is produced a lot of soldered connections which we assume are of satisfactory quality. The next step is to verify that this assumption is correct and this verification work should properly be classed as "inspection", since it is nothing but a measuring job and adds nothing to the quality of the work that has been performed.

When performed as a means of verification, inspection usually involves an examination of representative samples of product. In such cases two risks will be involved, due to the chance that the selected sample may not be sufficiently representative. In other words it may not contain the same percentage of defective units as the lot from which it was drawn. These risks may be considered as a consumer's risk and a producer's risk. The first represents the risk of accepting product of marginal quality and the latter the risk of rejecting product of aimed-at quality.

When sampling is used for quality protection, the two plans most commonly employed are "lot quality protection" and "average quality protection". Each involves rejection or acceptance of a lot of product depending upon whether the number of observed defects is greater or no greater than the "allowable" number for the particular sample selected.

Under proper conditions, sampling inspection is not only cheaper than detailed inspection but is usually more accurate and more informative. Determination as to whether or not conditions are appropriate for sampling and, if so, which type of sampling plan should be used, must include consideration of the following factors:

1. Are we interested in limiting the acceptance of more or less isolated lots of product to those lots which contain no more than a specified maximum percent defective? Or are we more interested in controlling the average percent defect of a series of lots of product which may later lose their identity in subsequent operations, in stockrooms, or in shipment?

2. What relation exists between the actual percent defective that results from our production process and the values we wish to use, either as limiting values for individual lot protection or for average outgoing quality protection? If the process average percent defective exceeds the limiting value within which we wish to control the average outgoing quality, or exceeds one half of the value established as the maximum for any lot, sampling is usually not economical. As the process average percent defective approaches these levels, we can no longer enjoy the benefits of small sample sizes and still retain a low producer's risk. This means that we are forced to spend more money, either in taking larger samples, or in fixing up the greater number of lots that will be rejected if we hold to a policy of continuing to take small samples under these conditions.

One of the major factors in keeping our process average percent defective at a low figure is, of course, the elimination of those causes of variability which are not due to chance alone.

Such items as abnormal variations in materials or machines, lack of skill, or carelessness in manual operations, rough handling, etc. are typical other-than-chance causes. Proper analysis of our inspection results will disclose the presence and magnitude of the effect of these other-than-chance causes. The more rapidly they can be eliminated, the more rapidly can we go back to small samples and enjoy the resulting economies of this form of inspection. By the same token, improvement in the process average percent defective, by the elimination of these other-than-chance causes, will result in less re-work, less scrap, and a correspondingly smaller unit production cost. It should be kept in mind, however, that there may be circumstances where even after the other-than-chance causes have been eliminated, the process average percent defective is still too high. This means that we have control, but the control is around the wrong level. The only thing to do under these circumstances is to change our facilities and processes, or to change our objective quality level.

RELIABILITY

Before we leave the subject of inspection, I would like to emphasize one more point, and that is reliability. Since successful application of quality control principles depends very largely on inspection and inspection results, our primary concern should be that the inspection results are as reliable as possible - whether they are used for protection or for informational purposes.

Here we are faced with a very serious problem. The only way we can secure entirely accurate information with respect to any given number of parts or things, is to inspect each of them under

conditions where there will be no error in measurements or judgment. The opportunities for operating inspections under such conditions are, of course, extremely limited. In practice we have indeterminate inaccuracies in the results of almost any 100 percent inspection, due to human error. On the other hand, the results of sample inspections may not always provide a true indication of quality owing to the laws of chance. Fortunately, however, in sampling inspections we can calculate the influence of chance, and the effect of the human error is likely to be considerably less because of the interest incentive associated with sampling inspections.

Another, and very important, factor influencing the reliability of inspection results, is the point of view of the inspector. Wherever possible, inspectors should be insulated from exposure to the day-by-day problem of production schedules and costs. They should try to put themselves, mentally, in the position of the recipient of the product - whether it be the customer or the production group which is going to use the product as a part or sub-assembly in the next operation. If any inspector allows himself to be influenced ever so slightly by a desire to contribute to the output or cost bogies of a production group, because he is one of the gang, he is likely to look less critically at marginal cases, and if at the end of the week or month output and cost objectives appear to be in jeopardy, he may also assume that things slightly outside of limits are also marginal, and that the shop or production unit should be given the benefit of doubt. In sampling inspections where there is an allowable number of defects for a given sample size, these emotions and influences may

result in an inspector regarding one over the permissible number of defects as the workings of an unkind fate which threw the extra defect into the particular sample he selected. This can very easily lead to the conviction that had he taken a different sample he would not have found one over the allowable number of defects and it should, therefore, be disregarded.

This dissertation on the reliability of inspection results is intended to highlight the effect of the point of view and the influences of emotion on the inspector rather than any question of honesty or dishonesty. It is important to recognize that unconscious inaccuracies can, and probably will, be present in inspection results if the inspector, or his immediate supervisor, has to meet cost or production schedules. Under such circumstances, it is too much to expect that inspectors can consistently retain the necessary unbiased point of view.

While I have included such expressions as "protection" and "control" under this discussion of inspection, I wish to make it crystal clear that inspection per se, is nothing but a measurement. It does not improve, protect or control quality. In fact too much inspection may well do the opposite. It is the action we taken on the basis of inspection results that improves or controls product quality.

ECONOMIC CONSIDERATIONS

What then do we get out of inspection? Is it worth what it costs?

In trying to answer these questions it will simplify matters if we deal first with material, piece part, sub-assembly and other process inspections. These may be divided into two categories for discussion:

First, there are those cases in which a definite minimum process inspection should be performed regardless of how good current quality may be. This is in the nature of quality insurance and is especially necessary for those process requirements which are not, and possibly cannot be, verified by inspection of the finished product. Many of us drive a car year in and year out without being involved in a single serious accident, yet we would not think of being without liability insurance because we know that if we ever were involved in one the results could be so seriously damaging that we refuse to run the risk. Likewise, if an important material or process got out of control without our knowing about it promptly, the adverse customer reaction that would probably result and the damage to ourselves in terms of reputation and replacement or repair costs would be so great that we should never go below a specified minimum inspection, how ever good current quality might be.

The kind and extent of this minimum inspection should not, therefore, be determined by how many defects we have found or have not found on past inspections, but by an estimate of the damage which could result if the process did get out of control.

The second class is much larger and includes those cases where something more than the basic minimum process inspection will prove to be worth while from an economic standpoint. Just how much more than this basic minimum inspection should be made depends upon actual quality, cost of the increased inspection and the value of the benefits gained. In general, manufacturing processes provide a certain amount of defective product which

results in scrap, repair and rework. The exact value of these losses will depend upon the amount of defective product and the particular stage at which it is found. Further losses will result from that portion of the defective product which inadvertently may be furnished to the customer because the defect is not readily detectable in the completed state. It should not be too difficult to equate the cost of protective process inspection against the savings in scrap, repairs, rework and consumer adverse reaction. An optimum point at which the cost of process inspection is no greater than the cost of scrap, rework, etc., can thus be arrived at and the appropriate character and amount of inspection prescribed. It is only by applying careful consideration of this kind to the development and specification of inspection procedures that producers can be reasonably sure they are meeting their quality objectives as economically as possible.

When it comes to inspection of finished products the dollar value return for each dollar spent on inspection is harder to identify and define. Against the cost of such inspection must be balanced the cost of complaint investigations, repairs or replacements which would result from selling unsatisfactory product, and also loss in reputation and good will. Some of these are rather difficult to evaluate but they are definitely in the picture.

Another advantage derived from reliable inspections is the probable improvement in producer - consumer relations, especially where consumers are willing to accept producer's inspection results as a partial or complete substitute for acceptance inspections that would otherwise have to be made

at the consumer's expense.

I should like to reiterate that in order to be reasonably sure we are getting our money's worth for inspection, the cost of each inspection operation should be weighed against the penalties of omitting the operation. These penalties can be considered as expense in terms of waste, abortive assembly, repair or replacement in our own plants, or adverse consumer reaction due to poor appearance, unsatisfactory operation, insufficient life, or any other failure to meet advertised claims.

CONTROL

Having specified our quality objectives and considered various methods of measuring quality, the next step comprises the establishment of procedures for controlling product quality at the desired level. In planning a quality control program, the following questions warrant very careful consideration:

> At what points in the process from raw materials through sub-assemblies to completed product should controls be introduced?
>
> For how many individual characteristics should separate control charts be maintained?

To answer questions of this kind requires sound engineering ability and experience, both with respect to the product characteristics and with respect to the production facilities and process. The program may be compared in its initial stages to preparing a questionnaire. Those of us who have at times attempted to get detailed information on the same subject from different sources, will realize that while the collection of the information may have taken a lot of

people a lot of time and effort, when we tried to piece it together we have found that we had not gotten exactly what we wanted. The fault may well have been our own, particularly if we did not spend enough time deciding exactly what it was we wanted before asking a lot of people to get it for us.

In setting up a quality control plan, we must then first decide exactly what kind of information will be most useful to us, in total and in detail. We can well afford to spend plenty of time on these considerations - ask ourselves, for example, how much will the information be worth when it is collected? What are we going to do with it? Which are the points in the manufacturing process which represent relatively large steps in the ladder of production costs from raw materials to the finished product?

These considerations will involve intimate knowledge of the quality characteristics which are necessary for production operations - i.e., hardness, fit of parts, etc. - and also those characteristics which are necessary in the final product to satisfy the consumer. Unless these facts are very carefully considered, there probably will be a temptation to cover more characteristics and introduce more control points and control charts than are economically warranted. Especially in those instances where management has not had very much experience with quality control techniques, and might be alarmed at the pretentiousness of a quality control program, it is well to start off with a small and modest plan.

One way to develop a keen management interest and appreciation of the value of quality control, is, of course, to have these techniques pay dividends. Fortunately, in a manufacturing plant where these techniques have not been used and management

may, therefore, be a little skeptical as to their worth, there is likely to be a virgin field of opportunity for quick and impressive savings. In such cases, it probably would be well to go after the cream and to aim quality control at some point where heavy scrap losses are occurring, or at a point which appears to be the focus of customer complaints. It becomes harder, of course, to continue to show new and important savings as we work our way through the cream and come down to the milk - but by that time we will have gained experience in how best to apply quality control techniques and can make more of these lesser opportunities. This is where the ingenuity of the quality control engineer is taxed most severely. It is, however, a part of quality control work which should attract high caliber engineers, and keep them interested.

I should like to emphasize that even with adequate quality specifications and adequate means of measuring quality, we cannot control product quality without taking action, and sometimes very drastic action, when our measurements show definite evidences of unsatisfactory trends in quality. In other words, it is no use spending the money to build a satisfactory alarm system unless one heeds the alarm when it goes off. Those who are unwilling to slow up, or even stop production, if necessary, in order to determine and correct the cause, or causes, of unsatisfactory trends will find, in the long run, that the cheapest thing to do with unsatisfactory product, is not to make it. Here is where a sympathetic management is essential since a quality control program has little chance of success unless it receives the wholehearted support of top management and all employees in the organization know that top management is

actively interested.

SUMMARY

In closing, I should like to summarize the items which, in my opinion, determine the degree of success of any quality control program.

1. The completeness and definiteness with which quality objectives are specified - for individual units of product and also for "lots" of product.

2. The cleverness with which control points are established and economic and efficient inspection plans are worked out for each control point.

3. The reliability of inspection results.

4. The ingenuity with which inspection results are analyzed and interpreted.

5. The promptness and vigor with which corrective action is taken when inspection results show the need for action.

NO. 3
Price 25¢

QUALITY CONTROL OF JET AIRCRAFT COMPONENTS

by

William C. Lee
Chief Inspector
Solar Aircraft Co.
Des Moines Plant
Des Moines, Iowa

for presentation at the

FOURTH NATIONAL CONVENTION
and
FIFTH MIDWEST CONFERENCE

of the

AMERICAN SOCIETY for Quality Control

June 1 and 2, 1950
Milwaukee Auditorium, Milwaukee, Wisconsin

FIFTH MIDWEST QUALITY CONTROL CONFERENCE
P. O. Box 1204, Milwaukee 1, Wisconsin

QUALITY CONTROL OF JET AIRCRAFT COMPONENTS

Applications of Practical Statistical Quality Control Methods in the Shop

By W. C. Lee*

This paper is devoted primarily to a description of practical shop methods for controlling quality during two types of manufacturing operations where conventional statistical procedures may not be easily applied.

1) Slow production items from continuous operations. (From 1 to 10 parts per hour)

2) Intermittent, short-run, low-quantity batch-type manufacturing.

Unlike high production items of average quality from continuous operations, where manufacturing conditions remain relatively constant, and a moderate quantity of completely rejectable parts can be tolerated, the production of small quantities of high quality aircraft material, at a low rate of production, under constantly changing conditions presents many problems and offers a real challenge to controlling quality by statistical methods.

It is doubtful if there are many manufacturing plants in the country who do not have some jobs of this nature and these are the ones that some companies may have hesitated to place under S.Q.C. In controlling the quality of stainless steel components for Jet Aircraft Engines there is no choice, since they are all relatively difficult from a statistical standpoint. We believe, however, that the procedures we have adopted for these

jobs may be useful in other metal working industries having similar problems. It is believed the illustrations to follow will prove that S.Q.C. need not be confined to high-production items alone and that it has a definite place in the low production field.

Solar Aircraft Company supplies many of the heat-resistant components for these engines, including the Ring & Outer Tube Assembly which house the Combustion Chambers. (Illustration A) One of the first parts to be put under S.Q.C. at the Des Moines Plant is the Ringholder component. (Illustration B) It is machined all over from stainless steel cast tubing (AMS 5645) at the rate of 6 per hour from each machine of a battery of 3 Gisholt Turret Lathes, in continuous operation. Tool mortality is high and tool changes are frequent. Tool and machine adjustments are necessarily numerous. Before applying Statistical Quality Control, scrap and rework costs were necessarily high and inspection time was correspondingly excessive. This was not an ordinary subject as the variables were numerous and their frequency high. (Illustration C)

There are several different methods for controlling quality in a manufacturing operation. One common method is to set up a Percent Defective Chart at an inspection bay and wait for the material to be inspected. Or we may inspect where the process is taking place. Sampling Inspection can be used of course by some approved plan and the lots which exceed the permissible percent defective may be screened and the defectives removed. After sufficient data has been accumulated so that the troublesome dimensions can be definitely identified, it is common practice to apply an Average and Range ($\bar{X}$ & R Chart) to the troublesome dimensions and attempt

to reduce or eliminate the assignable causes of defective material, maintaining the Percent Defective Chart to record the trend of the process and note any improvement.

After the trouble has been found and presumably eliminated it is common practice to abandon the X & R Charts and revert to the Percent Defective Chart completely, after the trouble has disappeared or has been reduced to an acceptable minimum.

We use this type of chart (Illustration D) wherever practical and it has been found very satisfactory, within its limitations. Note that the lower section of the sheet is a bonafide Percent Defective Chart and is plotted as such. The main columns list all of the characteristics to be inspected and the gages used for their inspection, comprising in effect, an inspection check list of the job. We record whether the material is undersize, oversize, reworkable, scrap or salvageable using the code shown on the chart. Since our inspection with a go and not-go gage tells us that it is undersize or oversize, we know whether it should be reworked, scrapped or used on a salvage tolerance. This of course facilitates disposal of defectives. Quality Control personnel are furnished with a list of permissible salvage tolerance to which we will accept off-print material (up to a given percentage.) Any material exceeding this salvage tolerance must be reworked or scrapped.

Left of the gage column is the inspection plan. Since our plant is under Air Force surveillance inspection, we lean heavily to double sampling per JAN-105. The Acceptable Quality Level is specified, designating the percentage of defective material which may be accepted on any group of

similar characteristics having the same A.Q.L. Note that the A.Q.L. varies from dimension to dimension. A critical dimension would receive careful 100% screening and the Acceptable Quality Level would be zero. Defects of a major nature affecting the malfunction, useability or interchangeability of a part would have a low A.Q.L. usually 1% or 2%. Minor discrepancies caused by workmanship only, such as finish, sharp edges, etc. are accepted up to 5% defective, the material being considered useable.

The purpose of this type of inspection procedure is to conform to Air Force requirements and to reduce stocking of <u>any</u> material which is <u>not useable</u> to a minimum. If the sample discloses <u>any</u> material beyond salvage tolerances the entire lot is 100% screened for that defect and the defectives removed for reoperation or scrapping. Unlike assembly methods by bolts and nuts, many of our components are welded to large assemblies and cannot be removed. We therefore cannot afford to stock unuseable material.

The total number of parts reworked, salvaged or scrapped may be individually totalized and recorded at the base of each column. The total number of defectives are entered in the appropriate block, the percent defective calculated and plotted as a regular P Chart. In totalizing the defects a part is considered only once in case two or more defects appear on the same part. For record purposes, the percent defective is estimated by dividing the number of defectives found in the <u>first</u> sample by the number of parts inspected in the <u>first</u> sample.

Over a period of time it is possible, by adding from left to right, to totalize the number of defects for any one characteristic, and identify the more troublesome dimensions. By the same reasoning, if a high plotting

appears on the P Chart it is possible to move upward and immediately identify the troublesome characteristic, and at the same time to obtain the trend of that characteristic over a period of time.

We have tried to incorporate in one form all of the information necessary for a preliminary study or a permanent record of any job. It is used in our Forming, Receiving Inspection and other manufacturing departments where inspection by attributes is applicable. Our Purchasing Department finds this type of chart to be of value in rating supplying vendors, in fact, vendors are supplied with a record of their performance by this method. Since a Percent Defective Chart records inspection by attributes, using fixed gages, it is an economical medium of recording material or process variation, particularly where tolerances are liberal and tool changes or adjustments are not too frequent.

The Percent Defective Chart has some limitations, however, when used on short-run or low-production jobs in a Machining Department where specifications are tight, tool mortality is high and the human element is a large factor, such as the following:

1) The P Chart does not provide a sensitive guide for the operator except to warn him of the quality of the material being produced. (We have found that the psychological effect of the $\bar{X}$ & R Chart on the operator is one of the most important factors in S.Q.C.)

2) It does not furnish a detailed record of tool wear, tool changes or machine adjustments. This is valuable data as it has a distinct bearing on the establishment of inspection frequencies.

3) It takes a greater number of pieces to discover which dimensions

3

are troublesome and by then the job is often well under way. While a P Chart will record the nature and direction of variations, it does not record their extent. On short run jobs it is important that this information be available as quickly as possible.

4) It provides information as to the proportion defective—it does not provide information as to how good or how bad.

So, while the Percent Defective Chart is a valuable economical tool for obtaining general information of a process, it has some limitations. It does tell us where to look for trouble and that of course is the first step toward eliminating it. It does not furnish the accurate information obtained by a thorough study, using $\bar{\bar{X}}$ & R Chart techniques.

Undoubtedly there are many different ways of controlling quality. The method we prefer to use and which we chose to illustrate here, while perhaps not as theoretically or mathematically advanced as some others coming into use, it works in the shop and is paying off in reduction of scrap and reoperation costs.

The procedure which we advocate for controlling quality on low-production short-run items is based on absolute and complete control of the process. This sounds imposing and expensive, but actually it need not be. It is simply a procedure for finding, recording and analyzing the causes of process variation, reducing the excessive variations to an acceptable level, then maintaining control of those variations at all times.

To be sure that this control is complete, it is first necessary to make a thorough study of the item to be placed under control. By this we

do not mean merely to determine the percentage of defectives by 100% or sampling inspection, but to determine and record not only the nature and direction but the extent and reason for every variation of every dimension during the study.

Without this type of study and control, we would never know the process capability of all dimensions in detail; how close to the borderline of rejection we were operating.

Most important of all, we need sufficient evidence to apply the minimum amount of inspection for a predetermined amount of protection, if the Quality Control program is to operate economically.

Furthermore, lacking a complete study, we might lose the benefits of valuable by-products which are nearly always obtained through a thorough study of a job.

After choosing the subject of our study, which is usually a high mortality item, the first step is to assign a competent Quality Control Engineer and a trained Process Inspector to the job. For several days every variation of every dimension is carefully recorded, particularly the reasons for abnormal variation. (Illustration E) All tool changes, tool adjustments and machine adjustments must be faithfully noted. Until we know how many parts a cutting tool will produce without a machine adjustment or a tool change we cannot accurately select our final sampling plan. No feed or speed changes should be permitted during the study period. (Note that space is provided for recording all tool changes, tool adjustments and major machine adjustments.) (Illustration F)

As each new bar of stock is put into use a recording is made on the chart. Set-up in the morning on cold machines is taken into consideration and the first few individual readings plotted as such. It is advisable to have 100% inspection for the first half-hour. About the third day sub-lot sizes and sample sizes can be determined and Sampling Inspection started on the less troublesome dimensions. As the process continues within the control limits established, Sampling Inspection can be gradually relaxed on the dimensions running well within control. However, more frequent inspections are maintained for the more troublesome dimensions.

Corrective action should be taken on these troublesome characteristics until the causes of excessive variation are eliminated. Corrective action can take many forms. Improved methods or tooling; different speeds or feeds of the machine; operator training; machine maintenance; investigation for consistency of material; specification changes; realistic rather than idealistic inspection standards; coolant control; improvements in previous operations; all with the thought of improving, not decreasing manufacturing efficiency.

When the study is completed and sufficient data accumulated, the control charts are removed and carefully analyzed by top Quality Control personnel. Where practical, dimensions having liberal tolerances and a known low range of machine variations are placed on a permanent Percent Defective Chart, maintaining an inspection after every tool change, tool adjustment or major machine adjustment. Average and Range ($\bar{X}$ & R) Charts can be held on a standby basis. They are retained on the most troublesome dimensions having larger process variations or tight specifications.

A minimum process-inspection plan can then be established and it becomes permanent.

Out of control plottings automatically require 100% screening of that periods' run and the largest and smallest individuals are considered for analysis purposes. Defectives are removed to a locked quarantine area. Reoperation of reworkable parts is performed before the set-up is torn down. Accurate records of reoperation are maintained. Prompt disposal of scrap is made to clear material records.

Operators should be trained to call for inspection immediately following tool changes or machine adjustments, and process inspection should be made available for this inspection. It is cheaper to provide this process inspection than to 100% screen material produced during the regular inspection interval. With this type of control very few defectives will be produced. A very small sample, taken at frequent intervals offers the greatest protection and gives the operator assurance that the process is satisfactory. These small but frequent samples are far more valuable to the operator than large samples taken once or twice a day. Assignable causes of discrepant material cannot easily be found for work produced hours ago.

Prior to applying S.Q.C. to this particular part a large percentage required reoperation for one reason or another, dimensionally or finish, before they could be presented for inspection. A thorough S.Q.C. study disclosed that all dimensional and finish requirements <u>could be met</u> with proper controls, and that final acceptance could be made at the machines. Reoperation has been reduced to a low figure. Final 100% inspection has

been completely eliminated, scrap losses are low and the parts go directly from the machines, through Zyglo Inspection into Stores.

Illustration F (of one dimension) shows final controls established and in operation. Nine charts are presently maintained on this job, however only one part per hour is inspected and we expect to reduce the number of charts as the job improves.

Since approximately 24 parts are produced every 4 hours, we could have chosen a lot size of 24 and a sample size of 4. We elected, for better control, to call our lot size 12 and our sample size 2, however one sample is inspected each hour and averaged each 2 hours for calculation purposes. This offers more assurance to the operator that the process is operating satisfactorily.

Every S.Q.C. study usually results in some interesting by-products. This particular study resulted in the release of a South Bend Engine Lathe and a heavy duty Drill Press for other work, this equipment having been used for performing reoperation on these parts. Material used for refinishing is no longer required. Handling has been reduced and special containers for these parts are more readily available. Large banks of parts are no longer necessary as they are useable when they leave the machines.

Manufacturing efficiency may suffer somewhat during a study and until final controls are established, however, in the final analysis a thorough study usually results in an increase in manufacturing efficiency.

The same type of control works equally well for intermittent, short-run jobs as for low-production, continuous operations. On short-run jobs

however, it is well to know the production schedule intimately, have all control charts ready and Process Inspectors alerted.

A study such as described reveals many conditions which contribute to excessive manufacturing time. Processing dimensions which receive subsequent machining are often held to unnecessarily tight tolerances and finish requirements. Dimensions which disappear in subsequent operations are often over-inspected. Illustration G shows the cross section of a part which was produced intermittently at the rate of two per hour. The same type of study was conducted as was described in the previous illustration. Among other things it was found that the operator was attempting to maintain a $\pm$ .005 tolerance and a 63 finish on the flange faces and outside diameter, and was making an excessive number of tool changes and tool adjustments. The Average and Range ($\bar{X}$ & R) Charts (Illustration H) shows tool wear and tool re-setting in attempting to hold the close dimensions and fine finish. (Illustration J)

The S.Q.C. study disclosed that these were processing dimensions which received subsequent machining and that the machining tolerance could be doubled to $\pm$ .010 and the finish increased from 63 to 250 micro inches. (Illustration H) indicates that tool changes and adjustments have considerably decreased as the result of these changes.

Manufacturing efficiency has increased on this item from 77.9% to 102.5%, part of which resulted from the study and the balance from by-products of the study.

Some of the greatest savings can be realized from by-products of S.Q.C. studies, often more than by control chart applications themselves.

3

We stumble over them every day and do not realize they exist until revealed as an afterthought to the original objective.

Every time a tool is changed to improve the finish of a machined part it must be handled several times. First it must be removed from the machine, delivered to tool grinding, ground, inspected, returned to tool stores, withdrawn by the operator and placed in the machine. Every time tool life can be increased it eliminates these steps of handling, regrinding time, and it reduces down-time of the machine producing the material.

Set-up control is very important. A set-up man or operator not having run a job for a week or a month, may make a few scrap parts until a correct setting is reached, unless set-up control is made effective. It is common practice for set-up scrap to be retained under a bench until a lot is completed perhaps indefinitely, resulting in hidden loss and inaccuracies of material control records. With effective set-up control, set-up scrap can be reduced to a minimum but it requires the coordinated efforts of the set-up man, foreman and process inspector. The inspector must be there when the <u>first</u> part is made and the machine should not run until it has been checked and the results known to the set-up man or the operator.

Segregation of material from controlled processes is also very important. We have established accumulation areas in each department and as material is accepted by the process inspector it is placed in these areas. (Illustration K.) When a sufficient quantity has been accumulated a Process Inspection Release is attached to the material and it may then be moved to the next operation or into Stores. No one is permitted to

3

place material in an accumulation area except a process inspector and no one but a Production Control employee may remove material from these areas.

Before each run is completed all defective material which has been held by the process inspector in a quarantine area, is reviewed by a Quality Control Engineer, the reworkable material is reworked before the set-up is torn down and the unuseable material is scrapped. This prevents odds and ends of defective material from accumulating and keeps stock records current.

While a great deal of our work at the Des Moines Plant consists of forming and machining of stainless steels nearly all of the completed units we produce are assembled by some form of welding. In the field of metallic-arc or acetylene-gas welding the greatest variable is the skill and consistency of the operator or the lack of it. In attempting to control these variables we are actually controlling the individual rather than mechanical variables.

Resistance welding on the other hand, contains many variables beyond the control of the operator such as electric power input; air pressure, water volume, water temperature, electronic controls; condition of electrodes and the variation of inspectors' opinions as to the acceptability of the weld.

Since the testing of resistance welding is necessarily destructive it is obviously impossible to 100% inspect the end product. Therefore the only method of assuring the quality of welded products is to control the process. Visual inspection of resistance welding except for external defects is wishful thinking. Industrial X-ray is commercially uneconomical

for such inspection. Sonic testing for this purpose is yet to be developed.

Inasmuch as our welded assemblies require maximum structural strength, to Air Force specifications, we must be able to guarantee the quality of all welded products in our production lines. We have adopted Statistical Quality Control methods on all resistance welding operations at the Des Moines Plant and the Resistance Welding Control Chart (Illustration L) shows the type of control which is maintained. Our Resistance Welding Inspectors are laboratory trained in all phases of physical testing and microscopic examination of welding, tensile testing, sectioning and micro-etching, mounting and polishing of specimens.

In respect to training, the entire supervisory staff of our plant have received approximately 15 hours basic training in S.Q.C. fundamentals and applications. Hourly paid inspectors receive approximately 10 hours basic training in the use and plotting of control charts.

The selection and control of Quality Control Personnel (Engineers and Statisticians) is important as their approach, manners and ability to make friends can materially aid a program.

Our approach to the labor unions was very open and the negotiating committee and all union stewards attended a condensed two hour discussion on the fundamentals and purpose of S.Q.C. All of this was accomplished before any control charts were applied in the shop.

We have found that it is wise to locate the first installation where supervision is strong and receptive to progressive methods. Properly presented and diplomatically executed S.Q.C. will sell itself to intelligent supervision, for they realize that it will make their jobs

easier and their efforts more effective. Factory supervision realize that every process inspector is another pair of eyes for the foreman. It is essential to gain the confidence of production supervision and to give them full credit for their contribution to the cause of Quality Control. It is never wise to claim that the Quality Control Engineer did the job.

It is very important to brief hourly paid operators before applying control charts to machines. Everyone is suspicious of something new that they do not understand. In this connection, machine operators were taken into a conference room and the principles and purposes of S.Q.C. were explained to them, and each worker was provided with a simple shop manual, supplementing the briefing. With this type of approach, grievances, because of S.Q.C. applications has been unknown at the Des Moines Plant. (Section of Manual shown in Illustration M).

The assistance and guidance a control chart provides for a machine operator is an asset to the Quality Control Engineer which should not be underestimated. The psychological effect on the operator to produce acceptable material is one of the most valuable tools of S.Q.C. The fairness of the control chart appeals to the average operator, if we can make him feel that the chart and the process inspector are there to help him produce better work.

To those who may be skeptical of the dependability of process control on slow-production or intermittent, short-run jobs, it may be said that the accuracy and dependability of this control depends to a great extent upon the thoroughness of its preparation and the manner in which it is executed by the Quality Control group, the process inspectors and super-

vision in general. It will be only as effective as it is prepared and executed. We could cover the walls with control charts but if the operators and the foremen, the men who produce this material, do not want them or use them effectively we have failed in our undertaking.

For those who would attempt to reduce the necessary time to thoroughly study a job, we may say there are no short cuts if the study is to be a success. Rushing the study or applying controls without sufficient data will only end in negligible results or mediocre success. Perhaps the study cannot be completed during the first run. It may be necessary to continue the study the following week or the following month, but once it is properly completed and the necessary information gathered, properly analyzed and adequate controls established, it should be good forever more unless some assignable cause or variation creeps into the process. But if controls are properly established discrepancies will be quickly discovered and stand-by controls in the form of Average and Range ($\bar{X}$ & R) Charts can be promptly re-established.

Perhaps there are faster or better ways to accomplish the results we have tried to achieve--later or more advanced techniques. Our main purpose here has been to illustrate one way of accomplishing these results in a manner that the operator and the shop foremen can understand, and the Plant Manager can interpret in dollar savings.

*Chief Inspector, Solar Aircraft Company, Des Moines Plant. Senior Member, Iowa Section of American Society for Quality Control.

3


ILLUSTRATION "A"

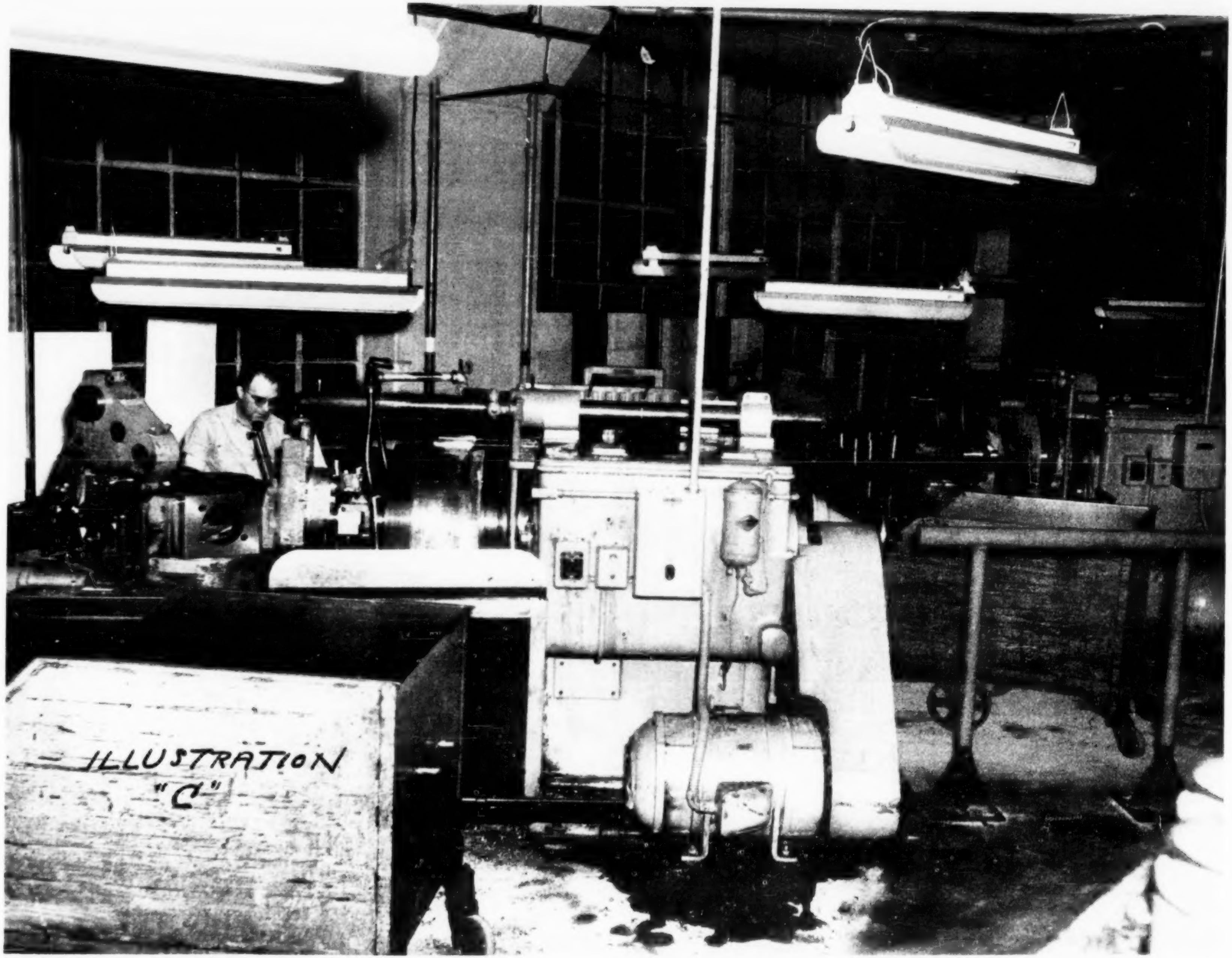


ILLUSTRATION
"C"

PERCENT DEFECTIVE CHART

SOLAR AIRCRAFT CO.
DES MOINES, IOWA

PART NAME RING HOLDER.

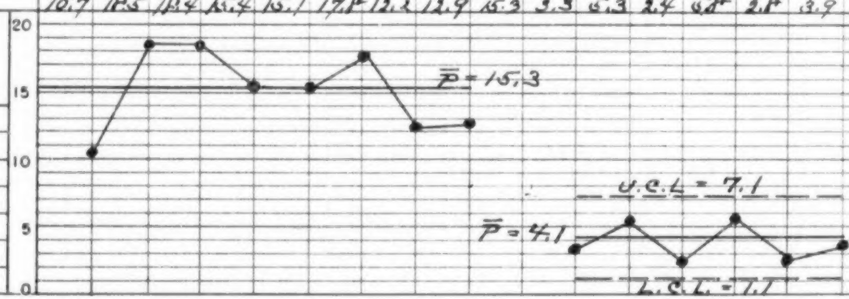
		1949						1950								
DATE & TIME		JULY	AUG	SEPT.	OCT	NOV	DEC.	JAN	FEB	TOTAL	3/10	3/17	3/24	3/31	4/7	4/14
INSPECTOR																
LOT SIZE		5189	2096	6577	5806	1381	3313	2453	2757	29573	392	386	243	430	349	406
SUB LOT SIZE																
LOT NO.																
SUB LOT NO.																

Class of defect	AQL	GAGES	CHARACTERISTIC	JULY	AUG	SEPT.	OCT	NOV	DEC.	JAN	FEB	TOTAL	3/10	3/17	3/24	3/31	4/7	4/14
				QUANTITY DEFECTIVE FOR EACH CHARACTERISTIC														
201	2%	VERNIER TAPE	7.625 DIA. ±.010											2-Δ		2-Δ		
301	3%	" "	7.188 " "															1-S
101	1%	6"-7" MICR.	7.150 " "										2-S 3+R	2-S 2-Δ				1-Δ
302	3%	VERNIER CAL.	7.037 DIA. ±.005										3+S		1+S		1+S	
401	4%	1"-2" MICR.	1.500 WIDTH ±.010											3-Δ	1-Δ	6-Δ		
402	4%	1" "	.875 " "											4+Δ				3+R
202	2%	1" "	.625 " "											3+R		4-Δ		3-Δ
102	1%	DIAL GAGE	.256-.259 GROOVE										3+Δ	4+Δ	2-R	5-R	3-R	4-R
303	3%	" "	.188 WIDTH ±.010												1-Δ	7-Δ		1-Δ
403	4%	SPECIAL GA.	20° X .173 CHAM.															
405	4%	" "	45° X .063 "															
501	5%	RAD. GAGE	.010 ±.005 RAD.															
502	5%	R.M.S. BLOCKS	63 FINISH												1R			3R
503	5%	" "	32 "													1R	6R	
504	5%	VISUAL	BURRS-SHARP EDGES															
TOTAL REOPERATION				501	345	1068	822	181	551	286	307	4061	5	3	3	6	9	10
TOTAL SCRAP				24	18	62	24	7	5	14	17	171	5	2	1	0	1	1
TOTAL SALVAGE				32	25	82	60	21	34	—	34	288	3	15	2	19	0	5
TOTAL NO. OF DEFECTIVES				557	388	1212	906	209	590	300	358	4520	13	20	6	25	10	16
PERCENT DEFECTIVE				10.7	18.5	18.4	15.4	15.1	17.8	12.2	12.9	15.3	3.3	5.3	2.4	5.8	2.8	3.9

TOTALS FOR 100% INSPECTION
DATA FROM PAST RECORDS.

CODE

R = REOPERATION
S = SCRAP
Δ = SALVAGE (MRB)
+ = OVERSIZE
− = UNDERSIZE



OPERATION NO. 10

DEPT NO. 17

INSPECTION PLAN 100% THRU FEB. 1949. UNDER S.Q.C. AFTER 3/10/50

ILLUSTRATION "D"

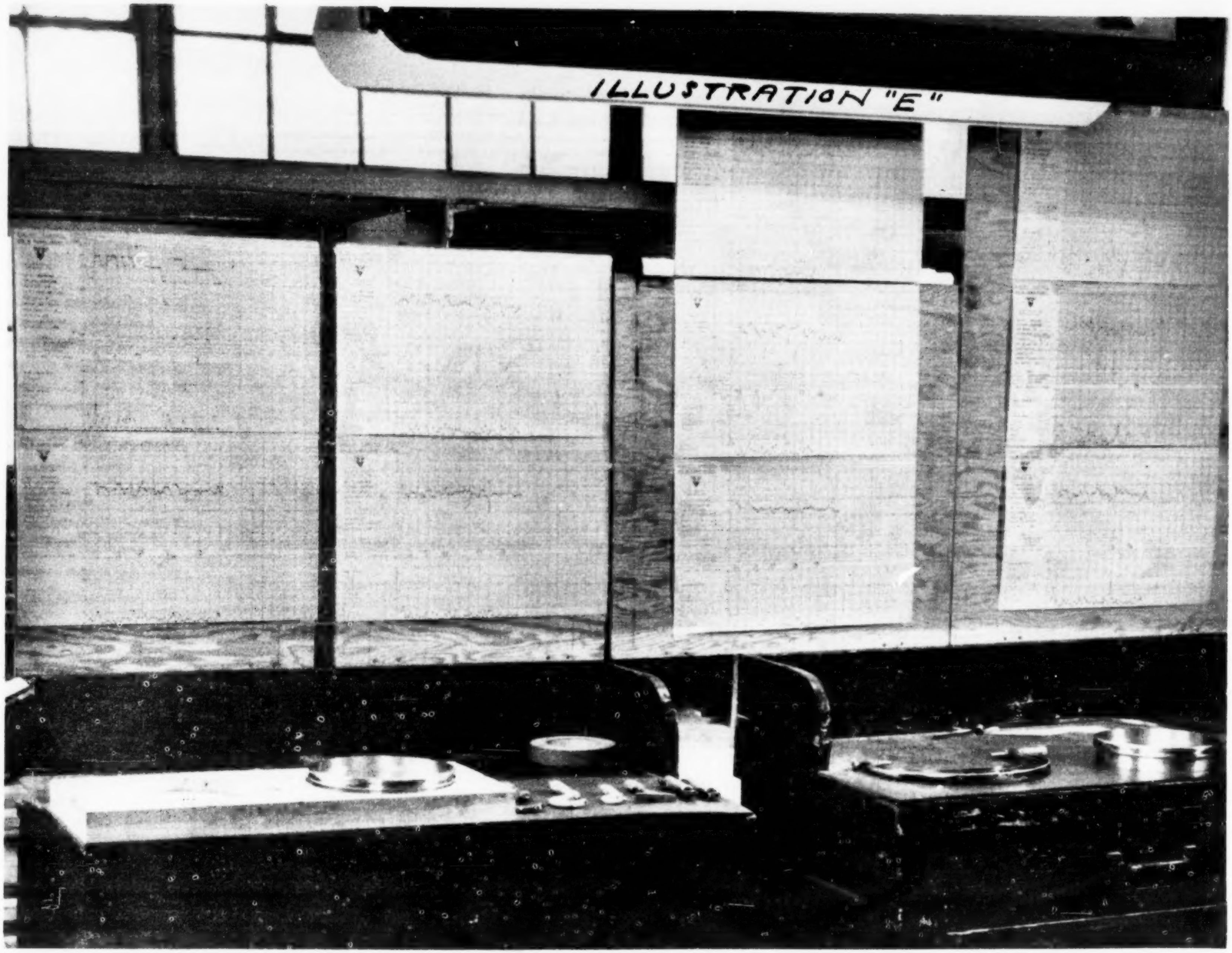


ILLUSTRATION "E"

QUALITY CONTROL
AVG & RANGE CHART

SOLAR AIRCRAFT CO.
DES MOINES, IOWA

PART NO. ______
PART NAME RINGHOLDER
ZERO SETTING 1.506
SPECIFICATIONS 1.496 - 1.516
OPERATION NO. 10
DEPT. NO. 17 MACH. NO. GISHOLT #1
HOURLY PRODUCTION 6
SAMPLE SIZE 2
FREQUENCY EACH 2 HOURS
CELL INTERVAL = .001

CHECK -
MACHINE ADJUSTMENTS
TOOL ADJUSTMENTS
TOOL CHANGES

RECORD - SPEED OR FEED CHANGES

ILLUSTRATION "F"

GRAND AVERAGE ($\bar{\bar{X}}$) = -.0018
$U.C.L._{\bar{X}}$ = +.0012
$L.C.L._{\bar{X}}$ = -.0048
AVERAGE RANGE ($\bar{R}$) = .0016
$U.C.L._R$ = .005
$L.C.L._R$ = 0
σ_x = .0014
$3\sigma_x$ = .0042
$6\sigma_x$ = .0084

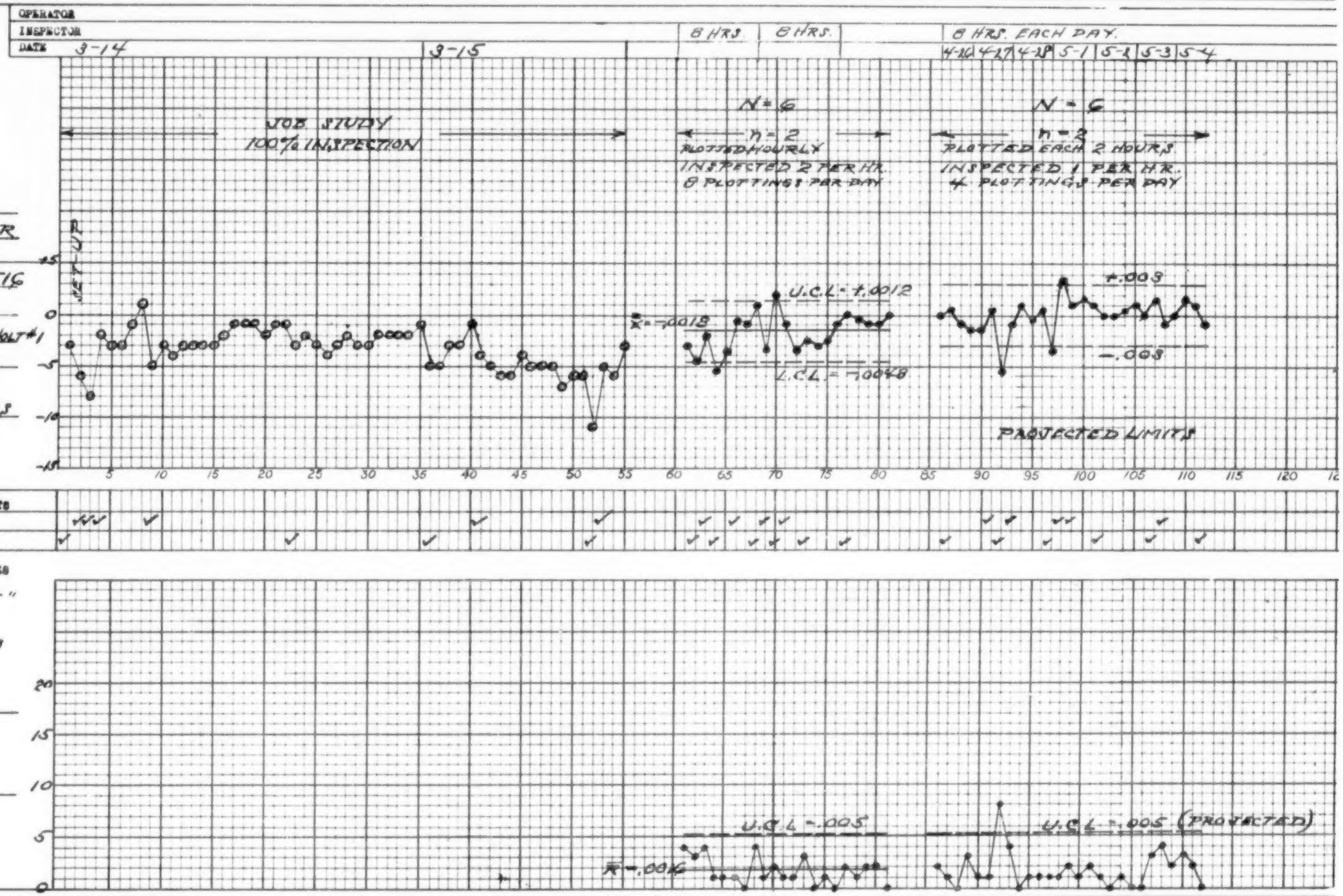
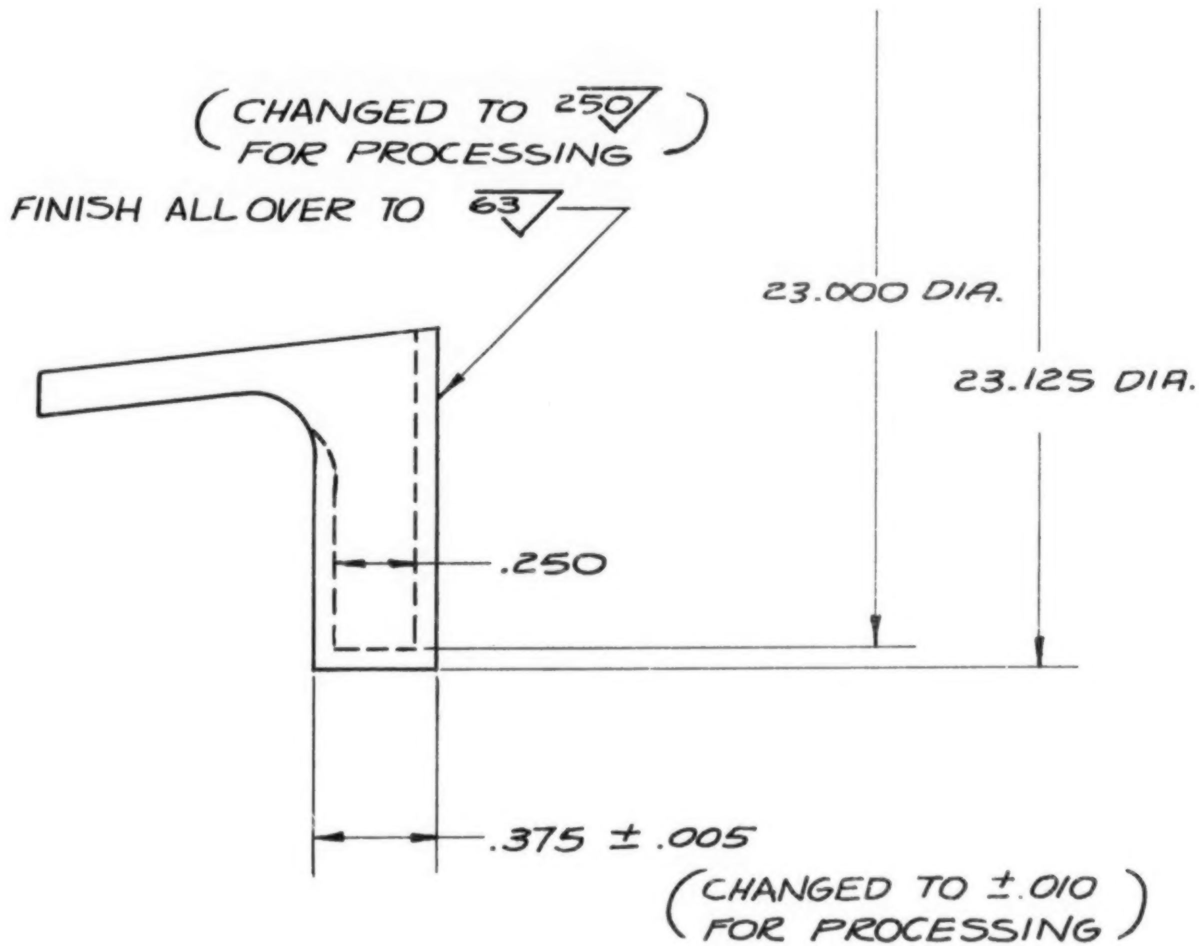


ILLUSTRATION "G"

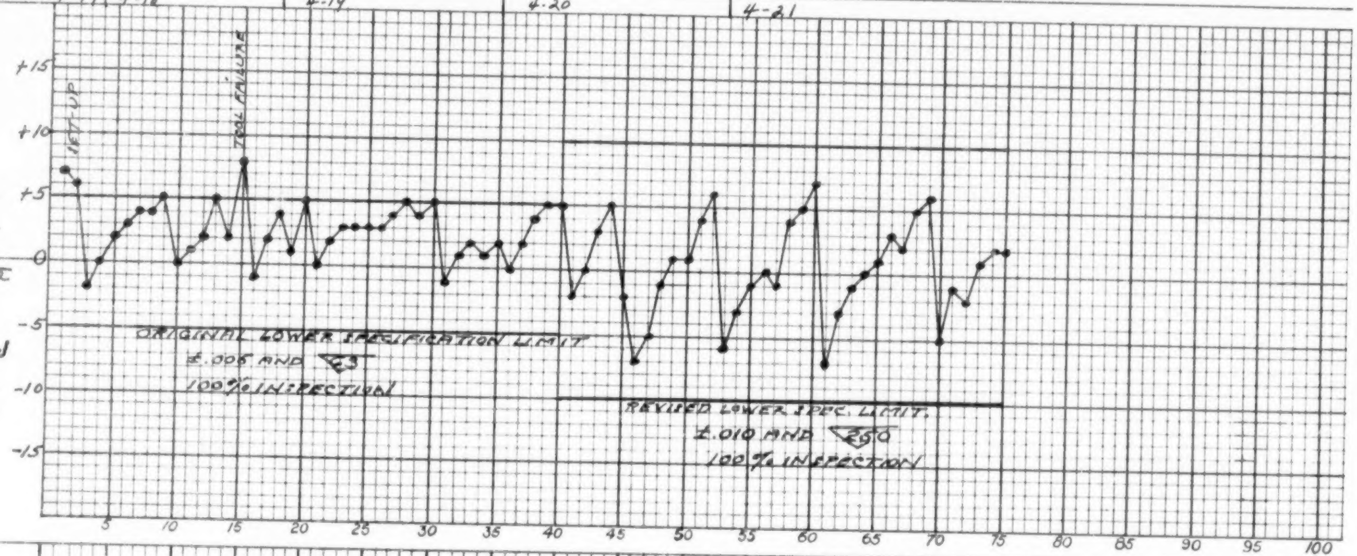


QUALITY CONTROL
AVG & RANGE CHART

SOLAR AIRCRAFT CO
DES MOINES, IOWA

OPERATOR 3375
INSPECTOR 2681
DATE 4-17 4-18 4-19 4-20 4-21

PART NO. ____
PART NAME FLANGE
ZERO SETTING .375
SPECIFICATIONS .370-.380 ∇32
OPERATION NO. 10
DEPT. NO. 17 MACH. NO. #3 P&J
HOURLY PRODUCTION 2
SAMPLE SIZE 100% INSPECTION
FREQUENCY ____
CELL INTERVAL = .001



CHECK -
MACHINE ADJUSTMENTS
TOOL ADJUSTMENTS
TOOL CHANGES

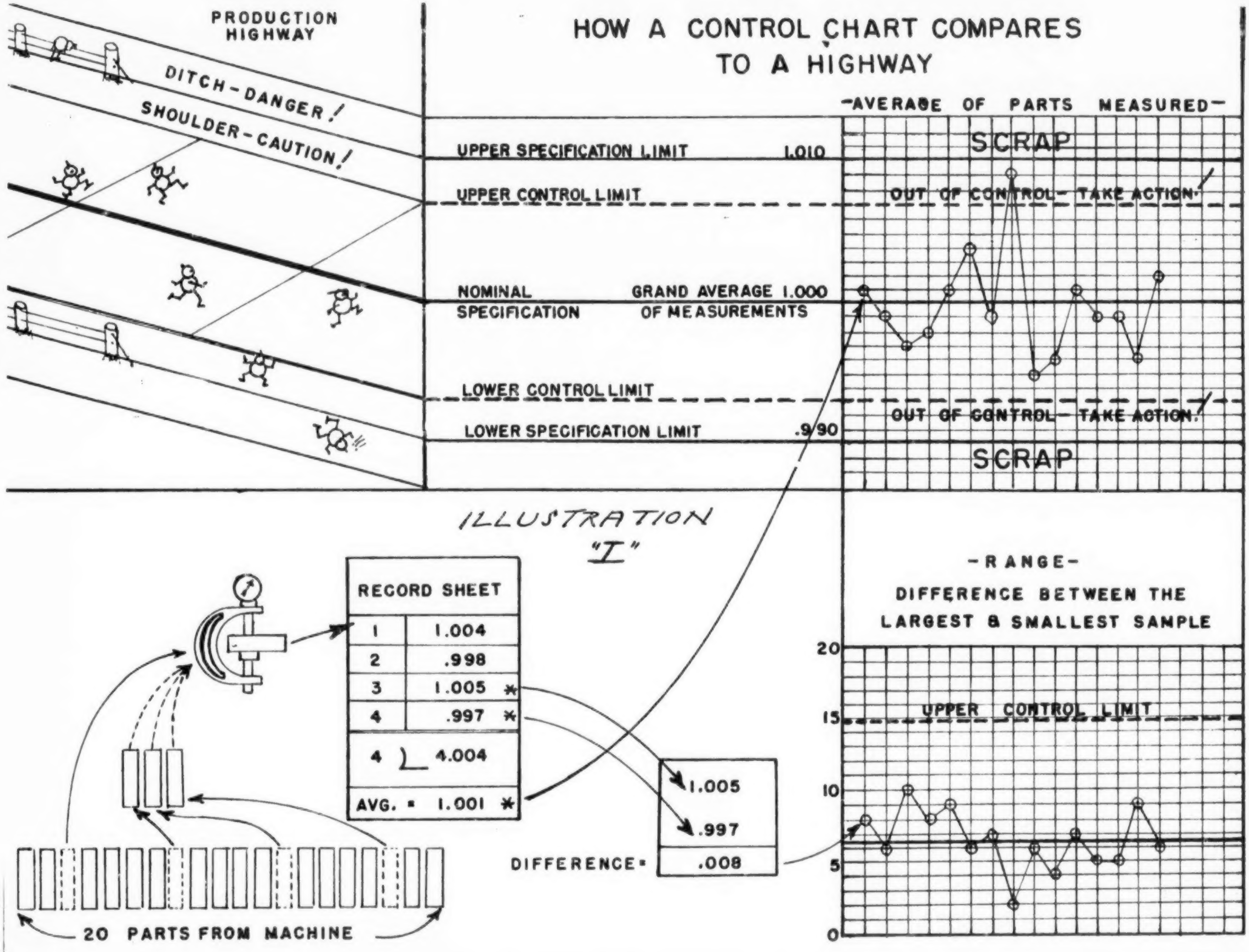
RECORD - SPEED OR FEED CHANGES

ILLUSTRATION "H"

GRAND AVERAGE ($\bar{\bar{X}}$) =
U.C.L.$_{\bar{X}}$ =
L.C.L.$_{\bar{X}}$ =

AVERAGE RANGE ($\bar{R}$) =
U.C.L.$_R$ =
L.C.L.$_R$ =

σ_x =
$3\sigma_x$ =
$6\sigma_x$ =



PRODUCTION HIGHWAY
DITCH - DANGER!
SHOULDER - CAUTION!
HOW A CONTROL CHART COMPARES TO A HIGHWAY
-AVERAGE OF PARTS MEASURED-
SCRAP
UPPER SPECIFICATION LIMIT 1.010
UPPER CONTROL LIMIT
OUT OF CONTROL - TAKE ACTION!
NOMINAL SPECIFICATION
GRAND AVERAGE 1.000 OF MEASUREMENTS
LOWER CONTROL LIMIT
OUT OF CONTROL - TAKE ACTION!
LOWER SPECIFICATION LIMIT .990
SCRAP
ILLUSTRATION "I"
RECORD SHEET
1 1.004
2 .998
3 1.005 *
4 .997 *
4 ⌊ 4.004
AVG. = 1.001 *
1.005
.997
DIFFERENCE = .008
-RANGE-
DIFFERENCE BETWEEN THE LARGEST & SMALLEST SAMPLE
20
15
UPPER CONTROL LIMIT
10
5
0
20 PARTS FROM MACHINE

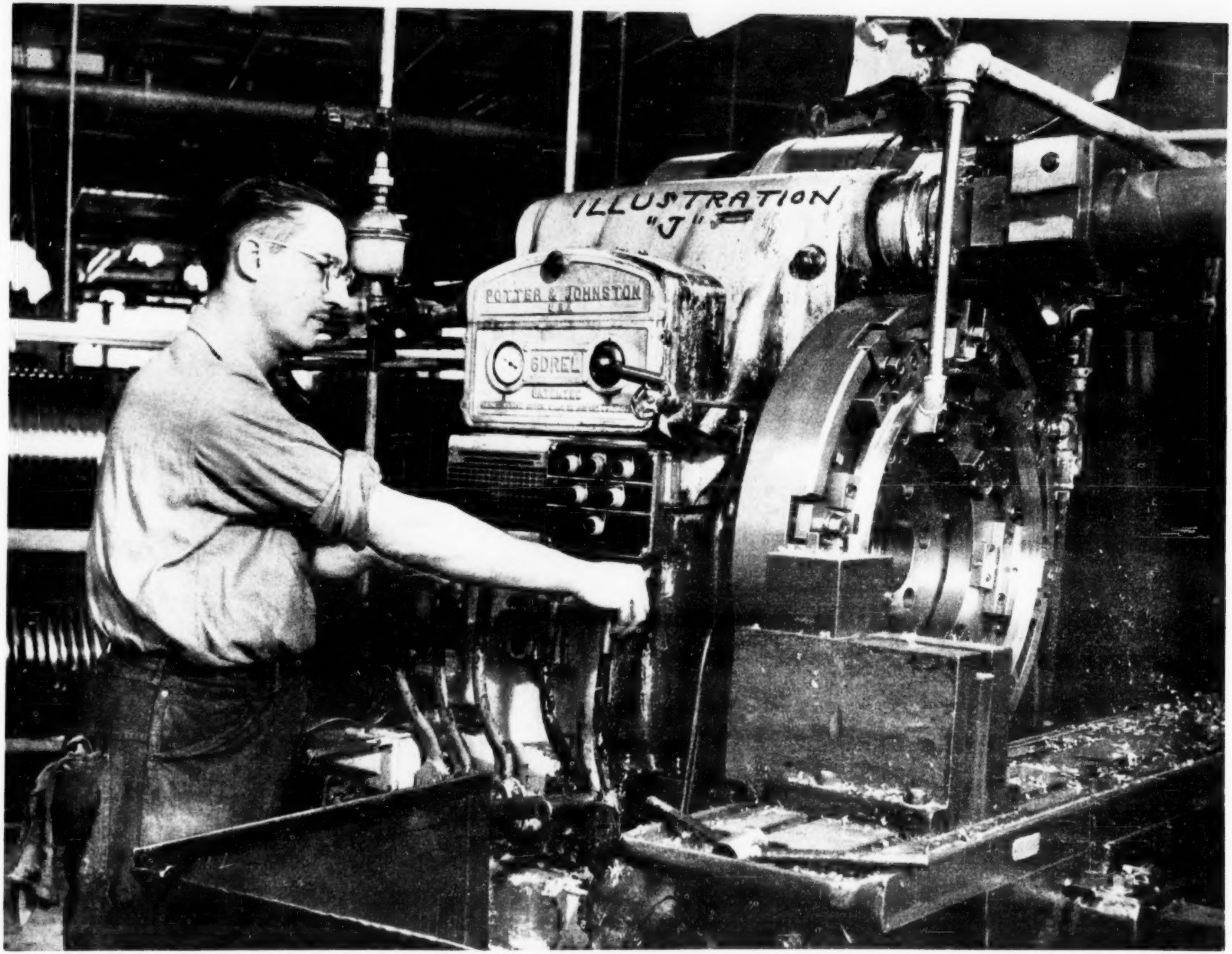


ILLUSTRATION "J"
POTTER & JOHNSTON
GDREL


ILLUSTRATION
"K"
QUALITY CONTROL
"ACCUMULATION AREA"
MATERIAL TO BE PLACED HERE ONLY BY
-PROCESS INSPECTION-
AND TO BE REMOVED ONLY BY
-PRODUCTION CONTROL-

RESISTANCE WELDING QUALITY CONTROL CHART

TYPE: SPOT FLOOR NO. 4 ASSEMBLY NO. W422 PART NAMES: REIN LOUVER TO SHELL MATERIAL: INCONEL TO INCONEL MATERIAL THICKNESS: .031 TO .043

WELDING MACH. NO.	SET-UP BY	INSPECTOR	DATE	TIME
2258	RJ	1113	3-3	800, 900, 1000, 1100, 1200, 100, 200, 300
2258	RJ	1113	3-6	800, 900, 1000, 1010, 1100, 1200, 100, 200, 300
2258	RJ	1113	3-7	800, 845, 920, 1000, 1100, 1200, 100, 200, 300
2258	RJ	1113	3-8	800, 900, 1000, 1100, 1200, 100, 200, 300
2258	RJ	1113	3-9	800, 810, 900, 1000, 1010, 1100, 1200, 100, 200

SPEED: ___ SPOTS OR INCHES PER MIN
TIME: 8 CYCLES ON ___ CYCLES OFF
TRANSFORMER POINT AND RANGE LOW
HOLD TIME
SQUEEZE TIME
OFF TIME
UPPER ELECTRODE NO. 1106-4
LOWER ELECTRODE NO. 1106-4

HEAT (1 LINE = 2 POINTS)
85
UPPER CONTROL LIMIT 80
NOMINAL SPEC. 75
LOWER CONTROL LIMIT 70
65

PRESSURE (1 LINE = 20 LB)
UPPER CONTROL LIMIT 1470
NOMINAL SPEC. 1400 ±5%
LOWER CONTROL LIMIT 1330
1300

SHEAR STRENGTH (1 LINE = 100 LB)
MIN. SPEC. LIMIT 840

FUSION (1 LINE = 10%)
100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 0
MAX SPEC.
UPPER CONTROL LIMIT 65
LOWER CONTROL LIMIT 35
MIN. SPEC.

NUGGET (1 LINE = 1/64 IN)
1/4 IN, 15/64 IN, 7/32 IN, 13/64 IN, 3/16 IN, 11/64 IN, 5/32 IN, 9/64 IN, 1/8 IN, 7/64 IN, 3/32 IN, 5/64 IN, 1/16 IN, 3/64 IN, 1/32 IN

ILLUSTRATION "L"

VISUAL
VOIDS
INTERNAL CRACKS
EXTERNAL CRACKS
INTERNAL POROSITY
EXTERNAL POROSITY
EXCESSIVE INDENTATION
EXCESSIVE METAL EXPULSION

COMMENTS-
NO WELD - PARTS HELD
LOW TENSILE - MACHINE STOPPED - DEFECTIVE TUBE - NO PARTS HELD
DEFECTIVE ELECTRODE - MACHINE STOPPED - 1 PART SCRAPPED. ELECTRODE REPLACED
DEFECTIVE ELECTRODE - REPLACED - INVESTIGATION STARTED - 1 PART SCRAPPED.

NO. 4
Price 25¢

STATISTICAL QUALITY CONTROL IN THE FOUNDRY

by

George Ver Beke
Quality Control Engineer
Union Malleable Iron Works of Deere & Co.
East Moline, Illinois

for presentation at the

FOURTH NATIONAL CONVENTION
and
FIFTH MIDWEST CONFERENCE

of the

AMERICAN SOCIETY for
Quality Control

June 1 and 2, 1950
Milwaukee Auditorium, Milwaukee, Wisconsin

FIFTH MIDWEST QUALITY CONTROL CONFERENCE
P. O. Box 1204, Milwaukee 1, Wisconsin

STATISTICAL QUALITY CONTROL IN THE FOUNDRY

By George Ver Beke,
Quality Control Engineer
Union Malleable Iron Works of Deere & Company, East Moline, Illinois
Senior Member, Iowa Section, A.S.Q.C.

I imagine we have had as difficult a time setting up a Quality Control Program as anyone. We started about five or six years ago by introducing Quality Control to two or three foremen in our plant by giving them a ten-day University course. After the completion of this ten-day course, they returned to their regular jobs of running a department. Naturally, these people did not have the time to spend to install Quality Control along with operating their departments. I assume that many of us have made the same mistake, treating Quality Control as a side line. We have found that to institute a Quality Control Program requires the full attention of whoever is installing the program.

About three years ago, it was decided to see if Quality Control, properly installed, would operate in a foundry. I was at that time approached on taking a ten-day course at the University of Iowa on the subject called "Quality Control by Statistical Methods". I just could not refuse a two weeks vacation with pay and a good expense account. After two days at Iowa City, I was completely confused. Why I ever thought this would be a two weeks' vacation, I don't know. After completing the course, I was very glad to return to work with my notebook full of notes which did not mean much to me. I took

my notebook and put it in my desk, hoping never to have anything to do with it again. About the second morning after returning to work, the general superintendent asked me what I thought of Quality Control. I could not let him know that he had wasted money by sending me to Iowa City. I told him that I believed Quality Control was a good tool and would do us a lot of good. Back in my own mind, I just could not see how a college professor could tell how to operate a foundry with less scrap by having a fraction defective chart on each machine. It is true that we both had the same interest at heart - less scrap - but I could not for the life of me understand how we were going to do this. Now, in our foundry we have a great many experienced operators - foremen who have been there thirty years, who have become experts on their particular jobs, and I did not think that I was capable of going out in the foundry and explaining to them how Quality Control was going to help them to reduce scrap. I knew that they would resent it. Actually, I would have, also.

We, as every other foundry, have tried different so-called black magics. When a salesman came in and told us he had a product to put in the sand to produce better castings with less scrap or produce better cores, we would naturally try some of his product. We would give it the fullest attention, and the results in the end would invariably be the same as we had before. We did have the satisfaction of knowing that we had tried to reduce our scrap. Now don't misunderstand - we still believe that in order to make advancements in the foundry industry, we should not be at a stand still. We should continue to try these so-called black magics. In the past, when trying these different products, the end result and effects derived would be strictly guess. Now, with Quality Control, we are able to prove whether or not we have advanced and how far. By using statistics, we are eliminating the guess work.

In July of 1948, I was called into the Manager's office. He approached me on installing Quality Control in our plant by saying: "Now that you have the belief and the education, I would like to have you install such a program." It took me about a week to decide whether or not I would undertake this project. After deciding to accept this position, I then went to the John Deere Waterloo Tractor Works. I wanted to see the best application of Quality Control available in our organization. At that time, their Quality Control Program had been in operation over three years. I found that the $\bar{X}$ R charts were being used to a very large extent along with the $\bar{P}$ charts. I returned with the zeal of an Evangelist, determined to prove that Quality Control would work at the Union Malleable.

We started our program by educating our inspectors and foremen, combined. Ernie Fay, who is the Quality Control Adviser of Deere & Company, was our instructor. He taught us the purpose of the charts and the benefits which may be derived from their use, touching on mathematics as little as possible. We did not want to frighten our inspectors and foremen by introducing them to a lot of theory. Mr. Fay also stressed the fact that we were to be of service to the foremen, and were not trying to take over their jobs or to run their departments. He also stressed the urgent need of cooperation between inspection and supervision.

When this twenty-hour course had been completed, I still did not know how we were going to sample hot castings to the extent where we could find much scrap.

It was decided to start in the core room. Good cores are one of the first steps to producing good castings. Our first chart was placed on the sandmuller, which mixes the sand for our coremakers. In the mixing of this sand, we use banksand, silica sand, corn flour, oil, kerosene, and water. We had a specification of 3.5% moisture. In making our study, we used an

$\bar{X}$ R chart, checking each half hour. We found that the moisture ranged from 1.5% to 6%. (Chart 1) We talked to our core room foreman as to what could be done to eliminate such a wide range. His answer was that, due to the bank-sand, we would never be able to do a better job of mixing without having a sand dryer, which would cost approximately $25,000.00. We then decided to check the moisture content of our banksand and found it to be 3% plus or minus 1%.

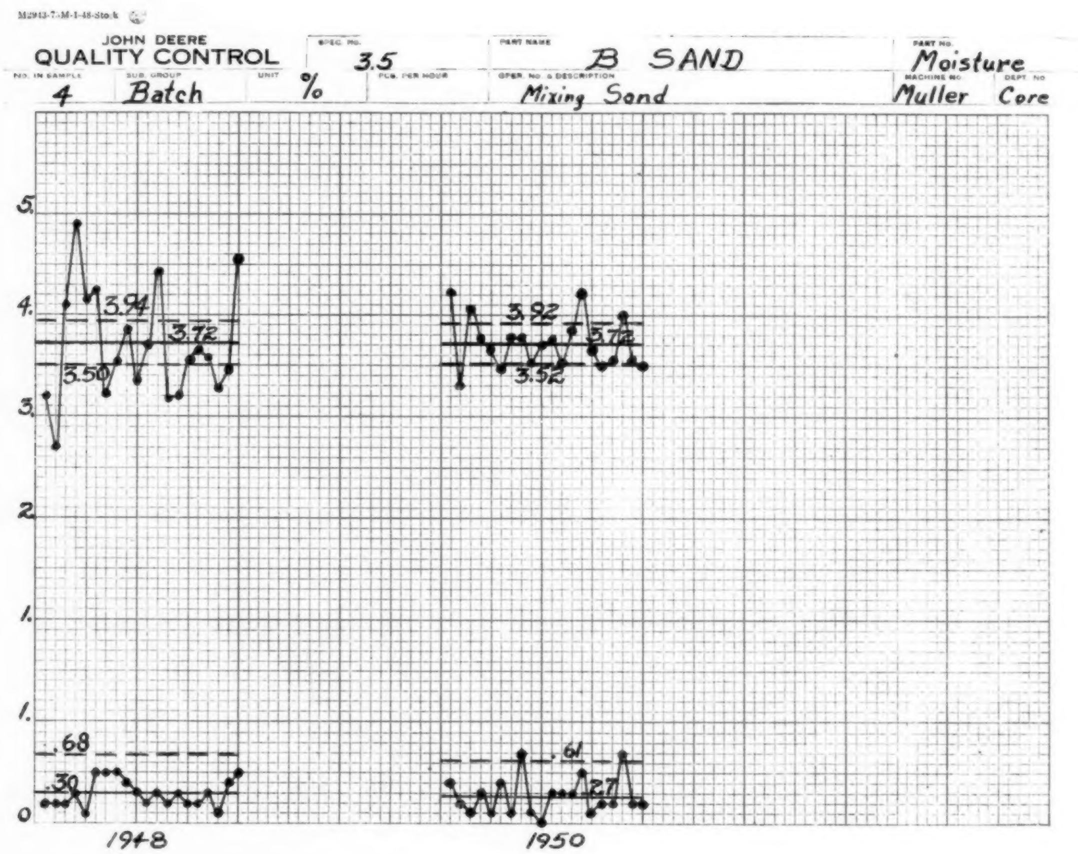


Chart 1 B Sand - Moisture

We were deeply concerned and started checking the sequence of operations and the equipment used. The first thing we found was that the water faucet dripped all during the period of mixing, continually adding moisture to the sand. We had the water faucet fixed. The next thing we found was a two quart container which was badly beaten up. I doubt if it actually held

one quart. It was lined with corn flour and oil, because we had been using the same container for measuring all the ingredients. We purchased new containers to be used for each ingredient. We then found after these corrections that our second shift operator could maintain a 3.5% moisture, plus or minus .7%. The operator on the first shift was still running 3.5% plus or minus 2%. This continued lack of cooperation resulted in a new operator. After about a week, we found the new operator operating a plus or minus .7%, proving to the foremen that this $25,000.00 sand dryer was not necessary to maintain proper core sand mix.

Just to prove that these coremakers could not tell the moisture in the sand, I will relate to you one experience which I had. This particular morning, a coremaker complained that his sand was too dry. We shoveled the sand out of the hopper, took it back to the muller, did not do a thing to it, and in about five minutes we returned the same sand to the coremaker, who began using it in about ten minutes. I went to this coremaker, asking him how the moisture of his sand was now. He stated if we continued to produce sand of this moisture content, he would never have any trouble making cores. Here is where Quality Control has taken the guess work out of the operation.

After completing our setup as to sand moisture, our next venture was in our core pack-off. Having no inspection previously at this point, we deemed it necessary to install Quality Control charts at this time. Our core pack-off consists of six women. These women take the cores off the plates after baking, throw out the scrap, remove the fins from the cores, and then put the cores into boxes. These cores are then stored in the boxes until needed in the foundry.

We started with a fraction defective chart, using a sample size of 20, checking each half hour. This was purely visual inspection. After a

couple of weeks, we found the quality of cores going to the foundry had greatly improved. (Chart 2)

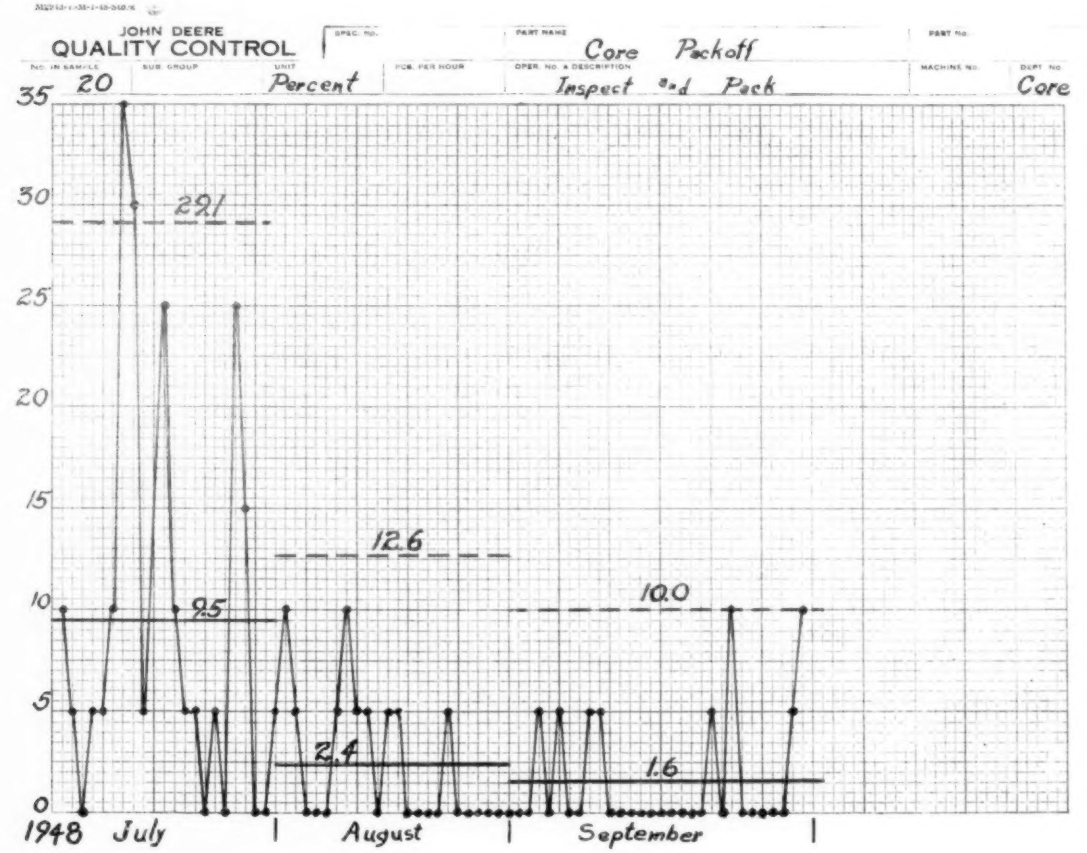


Chart 2 Core Pack-Off

We were being misled, because the foundry was continually running short of cores. This meant that one of two things was happening. Either the coremaker was turning in more cores than were actually being produced, or we were throwing away too many good cores at the pack-off. We checked at the pack-off on what these people were actually throwing out as scrap. We found at this point that it was here our shortage of cores was being created. What had happened is that these pack-off women had become so greatly interested in maintaining a desirable chart at 0% they were scrapping cores which should not have been scrapped. We then had to educate these

people on what was a defective core and what was a usable core. They were very cooperative, and immediately we knew we were on the right track. The difficulty of a core shortage in the foundry had been eliminated. The foundry foremen now often remark how much better the cores are as to quality, compared with cores previous to Quality Control. When, and if, we find faulty workmanship on any one check, we notify the foreman with a written reject notice. (Figure 1)

JOHN DEERE QUALITY CONTROL
REJECT NOTICE

PART NO.

PART NAME

MACHINE NO.

DEPT. NO.

REASON

NOTIFICATION	DATE AND TIME	BY INSPECTOR
1. FOREMAN		
2. FOREMAN		
3. GENERAL FOREMAN		
4. DIV. SUPT.		
5. PROD. SUPT.		

M2868-6-49 Stock

Figure 1 Reject Notice

Returning 30 minutes later, we check the same operator. If the condition causing the scrap on this job has not been remedied, we issue a second reject notice to the foreman, returning again in 30 minutes. If this defective material is still being produced or passed, a third reject notice is issued to the foundry superintendent. We, in turn, allow him one-half hour either to stop the job or correct the condition causing this scrap. If we find the same condition existing on the next check, a fourth reject

notice is written to the general superintendent. If steps are not taken to correct this condition in the next half hour, the fifth reject notice is issued to the general manager. To date, only six reject notices have reached the general manager. We use this reject notice system throughout the plant. At the beginning of our program, we issued 70 to 80 reject notices per day. As of now, we issue approximately 10 a day. And believe me, the foreman does not want any reject notices issued to the foundry superintendent. He makes sure to correct a job that is going bad at the time of notification.

After completing our setup on core pack-off, we decided to install Quality Control on our coremaking. Here again we had no inspection previously. We hired a man for inspection who had never seen the inside of the core room. He seemed very anxious to learn all about Quality Control. We trained him along the lines of inspection and what was a usable product. In about a month, we installed fraction defective charts on all our machines. We have approximately 25 coreblowing machines. Using a sample size of 20, again looking for visual defects, we found that the quality of cores was very poor. The coremakers began taking off scrap cores, produced a better quality core, and began to respect the Inspection Department. When we began our program, we had 6% of scrap being produced by the coremakers. This has been reduced to 1-1/2% as of now. There were several disagreements when we started our reject notices, as to whether the cores produced were of good quality, but after learning to live with each other and being able to decide correctly what was a usable core, our troubles ceased. We may now ask our core room foreman as to his opinion on Quality Control, and his answer would be "Don't ever take it out; it is helping me to do a better job much easier. I am now having less trouble with the foundry foremen and least of all with my superiors."

After completing the core room program, we were ready to make our first move toward installing this tool in actual foundry practice; that is, in the molding and pouring of iron. We have a completely mechanized foundry using a duplexing system in melting iron. Our metal is melted in a cupola and is brought up to temperature of approximately 2850° in an air furnace. We have two molding units, referred to as units 1 and 2, operating two shifts. Each unit is comprised of approximately 30 molders. Molding sand is conveyed to a large stationary hopper, located directly above each molding machine. The molds are placed on a roller conveyor, where they are held for pouring. We have six iron pourers continually pouring the molds produced by the molders.

An investigation of past records indicated that a very large percentage of our scrap was due to mis-runs. This investigation also revealed that No. 2 unit (day shift) was the greatest offender; therefore, we decided to make our first actual application in this division.

There are many different jobs run on this unit each day; also, many molds vary as to the number of castings produced from the mold. Therefore, our sampling plan could be only an experiment. We decided to check about five (5) molds each hour. Thus, the castings actually checked were five (5) times the number of castings in each mold.

Fraction defective charts were used for each individual job, and our first inspection consisted of looking for mis-runs and strains. The fraction defective in each sample was recorded on the chart and on the data sheet. We not only recorded the fraction defective, but the actual reason for the rejection.

We have taken the attitude that merely telling a foreman that he had 10% defective was not enough, but that we must tell him why the parts were defective, so he would know at once what corrective action should be

taken.

We are now experimenting along different lines to enable us to inspect adequate portions of the castings produced. We check each job at least once an hour. Whenever we find any pattern operating above 0% mis-run, we notify the foreman, giving him a sample of the defective casting. He, in turn, shows the casting to the iron pourer. The inspector returns one half hour later to check the same job, to see if it has been corrected. If not, we begin using our reject notices in the same sequence

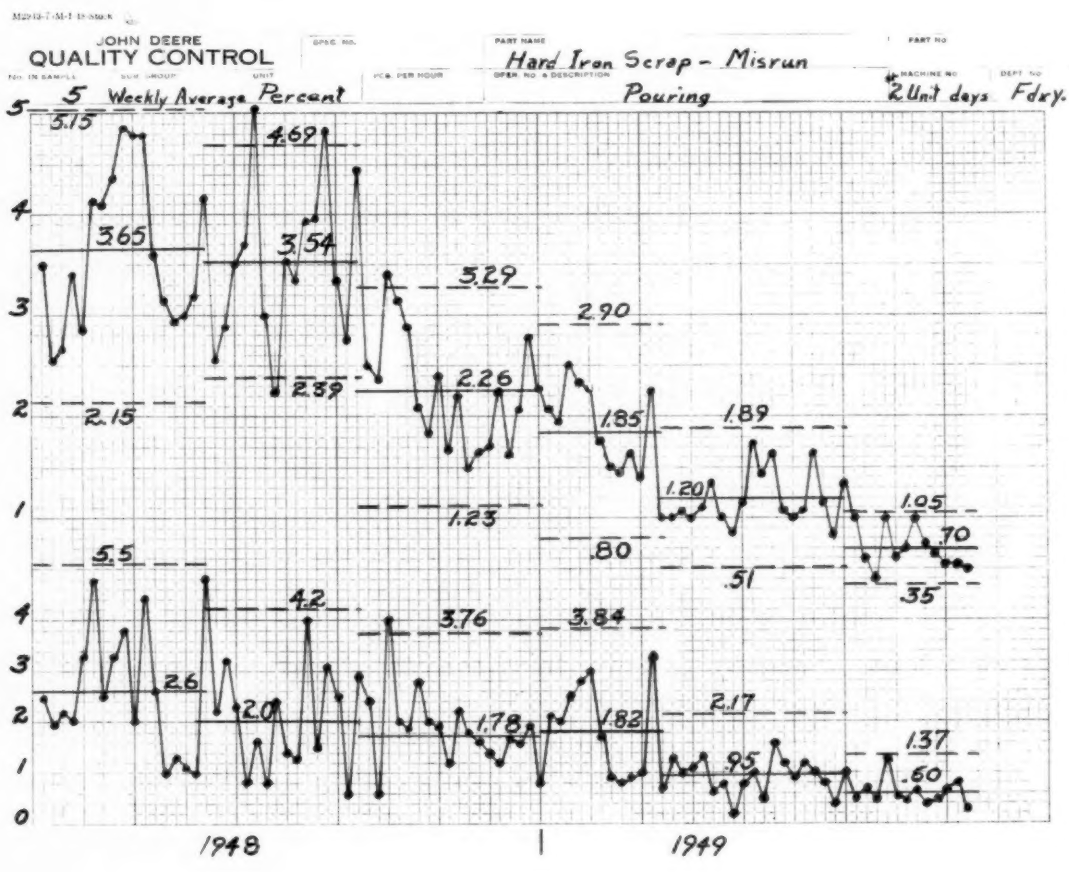


Chart 3 Hard Iron Scrap - Mis-run

In looking over our past scrap records, (Chart 3), our No. 2 unit

for mis-run, we found it to be 3.65%. Later this unit was operating at .70%, which proves that these professors did not have to know how to operate a foundry to enable them to educate us how to reduce scrap by using a P chart. We were all quite amazed by the results obtained by installing Quality Control on the pouring of iron. We had no idea that we could ever obtain a .70% mis-run week after week. After completing No. 2 unit, we installed the same method of inspection on our No. 1 unit, and it has proved equally satisfactory.

We were now ready to check for molders' defects, such as dirty, shifted, and strained castings. This is where we found our sampling to be a little difficult. Castings to be sampled are hot and dirty. We check our castings within 15 minutes after pouring. You may well realize the condition of these castings. It is rather difficult for an inspector to notice too many defects, which may occur unless he brushes the castings and has been trained and educated as to what defects actually are. We find it very important to stay in close contact with our Hard Iron Department. This is the department that removes the sprues and gates from the castings; consequently, this is the first department that will catch scrap if scrap has been produced. We realize that due to the undesirable condition of the castings in the foundry, we are bound to miss some of our defects which may occur, but by staying in close contact with the department which processes the castings immediately after cooling, we are, in many cases, able to help eliminate further production of scrap. Here again, we have a fraction defective chart using the sample sizes, as we did for mis-run, for each operator. Using the same reject system, we have reduced our molders' loss greatly. One reason for excessive scrap in our foundry was because, previously, we had a post-mortem inspection, which generally was a day or two late. The job was completed, or we had made two days of scrap without knowing it. We are now able to notify the foreman of scrap within 30 minutes of pouring. The

total time elapse from the making to the inspection is about 30 minutes. I think you will realize the importance of this sort of inspection setup.

In February 1949, we had completed our setup of Quality Control in the molding and pouring section of the foundry and were now ready to set up a similar program in our Hard Iron Department.

This department consists of about 35 people - two shifts - whose duties are to remove sprues and gates from the castings, chip off all core fins, throw out all scrap, and load the accepted castings into annealing pots. We also grind our lighter jobs while in the hard iron stage. We refer to these people as chippers and grinders. We have both men and women doing this sort of work.

Previously we had four inspectors who were referred to as "roving inspectors". The duties of these inspectors were more or less to police the department as to the quality of work. I now think a better name would have been "wondering inspectors", because all they actually did was wonder why the good Lord made the day so long. When it came to inspection, there was very little done.

We also had four scrap counters who were responsible for counting scrap castings thrown out by the chippers. Actually, all these people did was count the number of castings that had been delivered to them from the chippers, ignoring whether the castings were scrap or salvable. In other words, we were throwing away good castings. At this point, we started to absorb the people from normal inspection. We took two of these inspectors into the office and explained to them that they were to become Quality Control Inspectors.

In order to secure the full benefit from our new setup, we placed a fraction defective chart on each chipper and grinder. The inspectors would check a sample from a pot, to see that all the necessary operations

had been performed and if scrap castings were being thrown out. If one defective piece was found in the sample a second sample was taken immediately. If any defectives were found in this second sample, the pot was rejected and reclaimed by the operator. At the start 15 to 20 pots per day were being rejected. This resulted in several misunderstandings, and grievances were filed by the union, the operators feeling that we were being unfair to them. They did not realize that what actually was wanted was a thorough job of chipping with all scrap removed. It is true that it was necessary to retime some jobs, because of additional work that was added. However, this was a small percentage, and most jobs were timed to give us a quality product. At the present time, we seldom reject over one or two pots a day.

After the hard iron inspectors were trained and all machines charted, we decided a better job of inspection could be attained, by eliminating the four scrap counters, and by having inspectors count all the scrap at the time of our sampling inspection. This idea was not accepted with much enthusiasm by the inspectors, but after a little convincing and actual practice, the inspectors welcomed their new duties. This setup enabled the inspectors to see all the scrap; whenever salvable castings had been placed in the scrap hopper, the inspectors were able to return the castings to the operator for salvage immediately, thus eliminating unnecessary scrap. Another thing we found by having the inspector count the scrap at the time of operation was that we were able to notify the proper authorities on how much scrap we found on any particular job and for what reason. Consequently, we enabled them to prevent the same condition from continuing. We have one inspector, whose duty is to work between our hard iron inspectors and our foundry inspectors. This enables us to relay information missed in the foundry to the inspectors who are responsible for these particular jobs.

Just recently we have completed our program in the Finishing Department.

This department consists of four sections. The first section is the cleaning section. Here all castings are cleaned either by milling or sand-blasting after annealing. After cleaning, the castings are put into pans which hold approximately 2000 pounds. These pans are then weigh-counted and delivered to the grinding room. Here, again, we had roving inspectors along with scrap counters. We trained the normal inspectors in Quality Control, and placed P charts on each grinder. We check each hour to see that castings are properly ground and that the scrap is being thrown out. By having the inspectors count the scrap, we were able again to eliminate the scrap counters. The inspector is also responsible for checking each load before it leaves the department. He places a small tag which he has signed on each load, signifying that this load may be moved to the next department.

The third section of this department is called the bench room. Here castings are straightened either by a drop hammer, hydraulic press, or manual labor. A considerable amount of air chiseling of webs and drifting of holes is also done. The setup here is the same as in the first two sections. One inspector handles the department, releasing material by using the tag system.

The fourth section is our drill room. Here are located all drill presses, tapping machines, and punch presses. There are about 25 people working in this department. Checks are made on an hourly basis, and a P chart has been placed on each machine. Go and no-go gages are used on the majority of operations.

We are checking the operators for quality of work produced and for removal of all scrap from the load. Again, in checking our samples, if we find one defective, we take a second sample; if we find one or more defective pieces the second time, we reject the load for reclaiming by the operator. This inspector is responsible for all loads leaving his department, because this is the last department prior to shipping. He places a green tag on the

load if ready for shipment. A red tag is used if load is not properly completed or operations are missed.

About two months ago, we started to check all of our outgoing material by sampling each load to be shipped. We use the techniques of the Army Ordnance Tables as a basis for sampling to determine whether the load is to be accepted or rejected. We have noticed marked improvement in the quality of outgoing material. Our returns and complaints have decreased considerably. In one of our plants, our quality of work was running 5% defective. We checked with this plant about two weeks ago and found the percent defective less than 2.

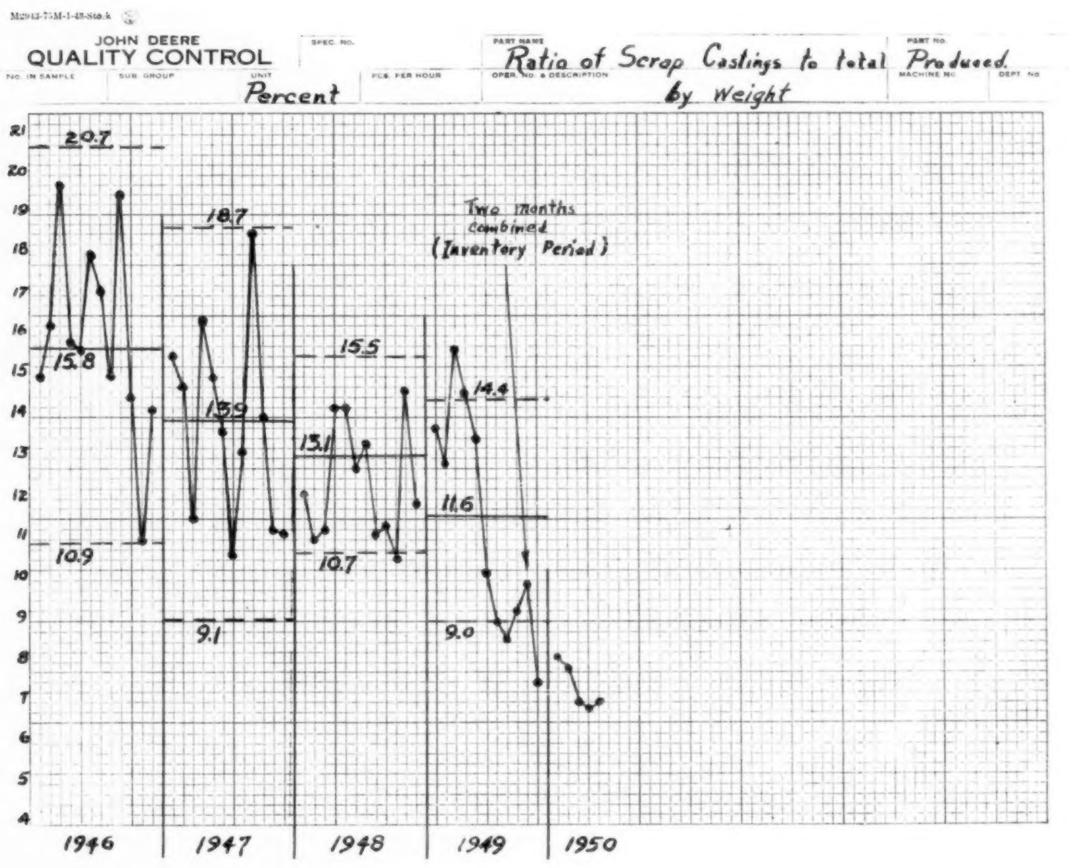


Chart 4 Ratio of Scrap Castings
to Total Produced

To point out the reduction in scrap, we refer to Chart 4, which gives the ratio of scrap to total produced, with the monthly and yearly averages from 1946 through part of 1950.

Since Quality Control was started in 1948, and time was given to take corrective action, and for operators to become familiar with the program, it will be noted that a continued downward trend in percentage of scrap has occurred; in the past few months less than 8%, whereas before Quality Control, the best year was 13%. Furthermore, as long as the trend continues downward there is a probability that the average will be still lower.

In order to operate the Quality Control Program, it was necessary to add two people to the inspection staff. However, this addition has made it possible for us to make inspections in the core room and foundry proper, where previously no inspections were made.

About five months ago, we decided that we were ready to apply Quality Control in the Vermilion Malleable Iron Works, another subsidiary of Deere & Company. This foundry is of a batch type, in which we have about 35 molds, who pour their own molds twice daily.

This foundry presented an entirely different problem than we had previously encountered. Therefore, considerable thought and study were required before we adopted any sampling methods, as we certainly did not want to upset any of the people, either workmen or supervisors, but we did want to get information relative to defects during the same day the castings were made, and not have to wait two or three days before we knew what quality was being produced.

We have four inspectors who check after each heat or time of pouring by dumping approximately 10% of the molds made by each molder, and then knocking off the sand and brushing the castings, so they are able to inspect these castings for various defects. After gathering this information, it

is plotted on a percent defective chart, which we call a molding machine chart. The inspector also carries and makes out an additional data sheet, on which he records the same information as is recorded on the chart and its data sheet. After all floors have been checked and all machine charts plotted, the additional data sheet is given to the foreman, who examines each individual job reported with any percent defective by an examination of the individual castings, with a view of determining the reason for the defect. The foreman then goes to each molder who has scrap and explains to him the percent defective found in his castings and the reason. This information is generally given within one hour after pouring.

After a few weeks, we found we had reduced the scrap considerably. Mis-runs seemed to account for about 40% of the scrap, or about 2% of the total castings made, and it indicated to us that we had not reduced this kind of defect very much.

We then picked out the 10 molders with the highest percent of reject due to mis-runs, and assigned two inspectors who were to check two molds out of each 7 or 8 molds at the time of pouring. Whenever one piece of mis-run was found in any sample, the inspector immediately notified the foreman, who, in turn, went to the molder and informed him of various means of eliminating mis-runs on that particular job.

We had found that at the beginning of a pouring period we had a greater amount of scrap than later on during the period, and investigation revealed that undoubtedly this was largely due to cold ladles.

After about two weeks' operation with the above plan, we found that we had reduced our mis-runs from 2% to approximately .75% and seemingly we are able to hold that level.

We are now in the process of installing pre-heating devices for our ladles, which, we believe, will assist us further in lowering our percent

4

defects per mis-runs.

In this smaller foundry, we have found conditions entirely different from those in the other foundries where we had installed Quality Control, in that a greater degree of cooperation already existed.

Prior to our investigation and the installation of Quality Control, this foundry was producing about 8% scrap, and after three months of information presented through Quality Control procedures, the scrap had reduced to less than 5%.

Reject notices are not needed in this foundry, because supervision takes immediate action as soon as they see a bad plotting, or as soon as the Quality Control man informs them of bad conditions.

As a further aid to keeping our supervision and operators informed, we have installed a large poster board, which displays each molder's name, together with his percent mis-run, miscellaneous, and total scrap. Also, on this board we have the rated position of each molder in the foundry, this being the actual standing so far as his percent scrap is concerned. We have found that these molders are very conscientious about their standing or rating and their own relative position with respect to the rest of the molders.

At the end of each week, we place a gold star after each of the five molders who have the smallest amount of scrap for the week, and this has proved quite effective, for it is obvious that no molder is very happy when his name appears at or near the bottom of the list.

The employees and supervision of this plant have accepted Quality Control as the useful tool and have shown splendid cooperation in the intent to produce better quality castings with less scrap.

4

NO. 5
Price 25¢

THE ROLE OF STATISTICAL QUALITY CONTROL IN AIR FORCE PROCUREMENT

by

M. H. Pardee, Captain
U. S. Air Force, Air Materiel Command
Wright-Patterson Air Force Base
Dayton, Ohio

for presentation at the

FOURTH NATIONAL CONVENTION
and
FIFTH MIDWEST CONFERENCE

of the

AMERICAN SOCIETY for Quality Control

June 1 and 2, 1950
Milwaukee Auditorium, Milwaukee, Wisconsin

FIFTH MIDWEST QUALITY CONTROL CONFERENCE
P. O. Box 1204, Milwaukee 1, Wisconsin

M.H.PARDEE, Captain, USAF
Hq., AMC, Wright-Patterson AF Base

ROLE OF STATISTICAL QUALITY CONTROL IN AIR FORCE PROCUREMENT

INTRODUCTION - The "Aircraft Nails"

In speaking of the "Role of Statistical Quality Control in Air Force Procurement" I feel that it will be more interesting to you, and, at the same time easier for me, if I build this discussion around a specific item of procurement which has long been a source of inspection difficulty, and use this example to illustrate those policies and procedures which collectively we call "Statistical Quality Control." The example that I intend to use is the case of the aircraft nail industry which, as far as the Air Force is concerned, is represented by five major producers. In order to respect their confidences, we will call them the Jack, Queen, King, Ace and Deuce Nail Companies. Now before any one asks what aircraft nails are I confess that the term "aircraft nail" is a pseudonym for an item which is manufactured in immense quantities and is of great importance to the Air Force. Except for the identification of the item itself, the facts which I shall present are completely authentic.

AMC Organization and Quality Control

Let me digress for a moment to present a brief picture of the overall organization within which we are trying to promote the use of statistical techniques. The details of organization are too extensive and too dry to discuss on this occasion. Let me just say that the Air Materiel Command is basically divided into three directorates; Supply and Maintenance, Research and Development, and Procurement and Industrial Planning. The Directorate of Procurement and Industrial Planning is itself composed of three Divisions; Procurement, Industrial Planning, and Quality Control; the latter concerns

itself with the evaluation of the quality of all material procured by the Air Force.

The AMC's Job - Its Scope

You can get some idea of the immensity of the AMC's procurement and inspection problems when I tell you that during the fiscal year 1949 the Air Force through the AMC procured about two billions of dollars worth of material of one sort or another, and does business with an average of 6500 contractors per year. In the New York procurement field office alone there are 1200 active contracts per month; speaking in rough averages 200 contracts expire and 200 contracts are instituted per month. It is clear that the Air Force exercises a profound effect on thousands of businesses and tens of thousands of people. Obviously, then, when the numbers with which we are dealing are so large, it is imperative that the Government in both its inspection and acceptance functions be fair to its vendors as well as to itself.

Surveillance Inspection Policy

Statistical Quality Control is, of course, concerned with developing and applying procedures to achieve such fairness. But before we discuss the details of SQC it is necessary to mention briefly the Quality Control Division's Surveillance Inspection Policy. It is basic to our discussion. By "Surveillance Inspection" we simply mean that inspection is actually performed by contractors at the direction of the Government. In other words, instead of attempting to perform inspection itself, the Government assures itself that the contractors' inspection system is adequate. This assurance is based on a thorough inspection of the contractor's methods, tools, gages, organization, and so forth. If such is the case, it follows that the inspection performed by this system will be satisfactory provided that the system itself remains

unchanged, a condition which the Air Force inspector assures by periodic follow up surveys. In short, we can say that the Air Force chooses to inspect not the material it is receiving but rather the inspection system the contractor has provided to insure that only acceptable products are delivered. This policy is based largely on practical necessity. The most casual investigation will reveal that it is manifestly impossible for the 1400 odd Air Force inspectors to inspect physically all the material that the Air Force procures. By the same token, if the Air Force were to hire enough inspectors to do that sort of a job, the cost would not only be prohibitive, but since it would require an almost exact duplication of of the contractor's inspection system; it would be grossly inefficient. But enough of this background discussion - let us return to the case of the nail industry. Incidentally throughout this paper I am presuming that contracts for purchases are let on the basis of competitive bidding.

Aircraft Nails - Quality Low - Inspection Costs High

Approximately six months ago the Statistical Quality Control Unit, which was just becoming operational, was asked to review the acceptance inspection procedures for (and the quality of) aircraft nails. A very comprehensive study indicated that the entire situation was unsatisfactory both to the Air Force and to the contractors. It was quite apparent that the Air Force was receiving inadequate inspection protection because the specification for nails was quite flimsy both because of what was and was not required. At the same time the contractors were unhappy for two reasons; first, quality evaluation methods were not standard and thus acceptance of material depended as much on subjective as objective criteria;

(this, of course, placed some contractors - even the best - in a less favorable competitive position than others). Secondly, the cost of inspection was excessive; in fact it cost more to inspect the nails than to manufacture them.

The Specification

Now, having discovered that there was much to be done, we felt that the specification was the major root of all this evil. So lets take a look at a portion of the applicable specification (Figure I) and study the acceptance requirements for dimensional characteristics only.

Extract of Acceptance Procedure for "Aircraft Nails"

F-5. Sampling.-

F-5a. Finish and Dimensions.- Sampling for finish and dimensions shall be random in accordance with Table VI. The acceptance-rejection provisions of Table VI only shall apply separately to each of the inspection elements or requirements. For example in the case of dimensions, two of the inspection elements are over-all length and shank diameter.

TABLE VI
Sampling for Finish and Dimensions

Lot Size	Sample	Sample Size	Combined Samples: Acceptance No.	Combined Samples: Rejection No.
Under 25	First	All		
26-75	First	25	0	1
76-500	First	25	0	3
	Second	50	2	3
501-5000	First	50	1	4
	Second	100	3	4
5001-20,000	First	75	1	6
	Second	150	5	6
Over 20,000	First	100	2	6
	Second	200	5	6

FIGURE I

Let us suppose we are nail manufacturers and that we are thoroughly scrutinizing the specification to determine to what quality standards our nails must conform to be acceptable to the Air Force. The fact that a sampling plan is written into the specification is in itself an important fact; the Government indicates thereby that it does not require evidence that the nails are perfect or even near perfect. So now we ask - To what degree may nails be nonconforming and still be acceptable? Paragraph F-5a,(Figure I)of the specification partially answers this question. You will note that this paragraph says that Table VI applies separately to each inspection element. That word "separately" is extremely important. A question may now be proposed "What percentage of nails may be too long and still be acceptable?" Lets look at Table VI and see what that tells us. Assume that our nails are made in three different lot sizes - N = 400, N = 4000, N = 25,000. We want to know what are the chances that a series of lots of any given size will be accepted if the percent defective for length is 1%, 2%, 5%, etc. By preparing operating characteristic curves for each of the sampling plans in Table VI we can estimate our chances of acceptance (Pa) for any given percent defective (p). The operating characteristic curves (Figure II) are very revealing. I need not comment on them in detail but you will notice that for the two larger lot sizes material approximately 1% defective is acceptable 95% of the time; however the smaller lot (76 - 100) are accepted only 75% of the time at 1% defective. Of course, the knowing manufacturer will waste no time in discovering that the lot size has an important bearing on acceptance. You can ponder the numerous implications - economic or otherwise - of these facts. But lets not stop here. Suppose the nail has 10 dimensions. What then? Does this mean that you could accept nails as bad as 10% defective? A survey made by the AMC has shown that the quality of the nails accepted by the Air Force reflects

COMPARISON OF CHARACTERISTIC CURVES FOR THREE SAMPLING PLANS

PLAN NO. 1	PLAN NO. 2	PLAN NO. 3
N = 26 - 75	N = 76 - 500	N = 501 - 5000
n = 25	n_1 = 25	n_1 = 50
c = 0	c_1 = 0	c_1 = 1
	n_2 = 50	n_2 = 100
	c_2 = 2	c_2 = 3

WHERE: N = LOT SIZE; n_1 = FIRST SAMPLE SIZE; n_2 = SECOND SAMPLE SIZE; C_1 = ACCEPTANCE NUMBER FOR FIRST SAMPLE; C_2 = ACCEPTANCE NUMBER FOR TOTAL SAMPLE

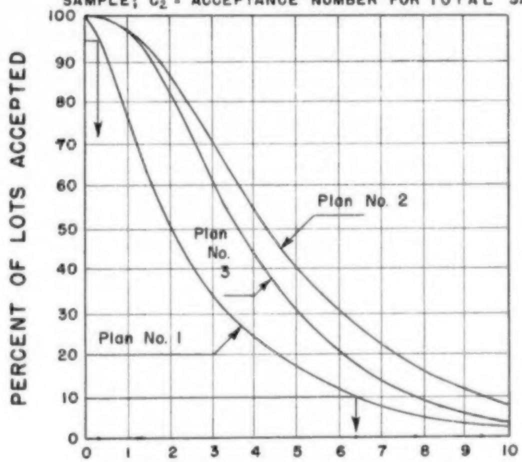


FIG II

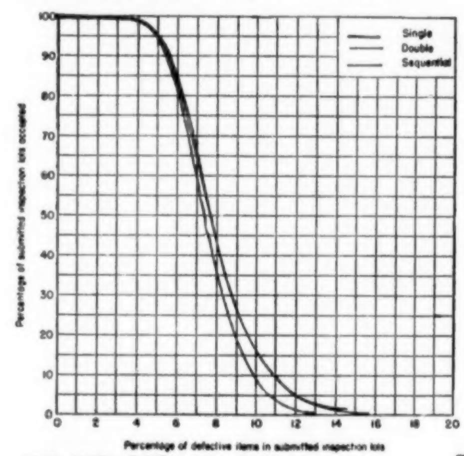
NHS 27-APR-50
QC. DIV. HQ-AMC

NORMAL INSPECTION — LOT SIZE 3,200 - 8,000 — AQL 5.0%

a. SAMPLING PLANS.

Type of sampling	Sample	Sample size	Combined samples: Size	Combined samples: Acceptance number	Combined samples: Rejection number
Single	First	225	225	17	18
Double	First	150	150	11	29
	Second	300	450	28	29
Sequential	First	50	50	2	7
	Second	50	100	4	11
	Third	50	150	8	15
	Fourth	50	200	12	19
	Fifth	50	250	16	22
	Sixth	50	300	21	26
	Seventh	50	350	25	29
	Eighth	50	400	28	29

b. OPERATING CHARACTERISTIC CURVES.



From JAN-STD-105

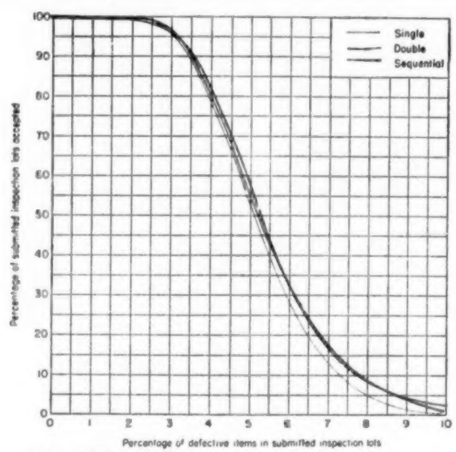
Fig. III

TIGHTENED INSPECTION   LOT SIZE 3,200 - 8,000

a. SAMPLING PLANS.

Type of sampling	Sample	Sample size	Combined samples		
			Size	Acceptance number	Rejection number
Single	First	225	225	11	12
Double	First Second	150 300	150 450	7 18	19 19
Sequential	First Second Third Fourth Fifth Sixth Seventh Eighth	50 50 50 50 50 50 50 50	50 100 150 200 250 300 350 400	0 3 5 8 10 13 15 19	5 8 11 13 15 18 20 20

b. OPERATING CHARACTERISTIC CURVES.



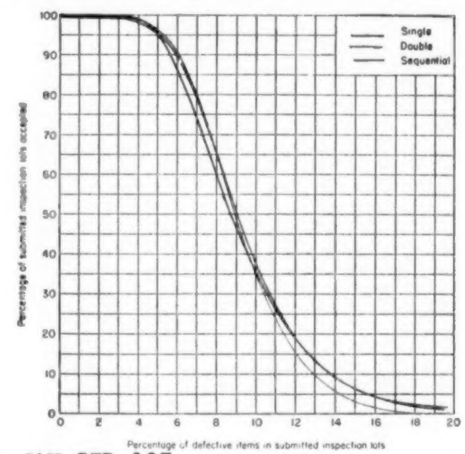
From JAN-STD-105   Fig. IV

REDUCED INSPECTION   LOT SIZE 3,200 - 8,000

a. SAMPLING PLANS.

Type of sampling	Sample	Sample size	Combined samples		
			Size	Acceptance number	Rejection number
Single	First	115	115	10	11
Double	First Second	75 150	75 225	6 14	15 15
Sequential	First Second Third Fourth Fifth Sixth Seventh	30 30 30 30 30 30 30	30 60 90 120 150 180 210	1 3 6 9 11 14 17	5 7 10 13 15 18 18

b. OPERATING CHARACTERISTIC CURVES.



From JAN-STD-105   Fig. V

the fact that the specification allows a rather liberal AQL for each characteristic.

Revision of the Specification

Obviously, if we are to do anything to improve the situation we must first direct our attention to the specification and define as clearly as possible the procedures by which material is judged to be acceptable or non-acceptable. This is, of course, a sizeable task because a sound inspection program cuts across both inspection by attributes and by variables, as well as numerous problems in sampling for various physical tests, destructive or otherwise. However to simplify this discussion let us restrict the scope of our problem to revising the specification only for inspection by attributes, i.e., for finish and dimensions, as mentioned earlier. Fortunately the Military Establishment has a standard which provides numerous sampling plans of considerable flexibility which can be used for inspecting nails and with considerably greater effectiveness than the plan we mentioned previously. This is JAN-STD-105; it is mandatory for use by the Armed Services for sampling inspection by attributes. Therefore the first measure we will take to improve the present inspection specifications for nails is to reference JAN-STD-105. This Standard is exclusively a group of Sampling Tables so, for a moment, let us see what happens when Ace Company presents for acceptance a series of lots varying in size from 3200 to 8000 units each. Assume that these lots are uniformly 5% defective. What fate awaits Ace's nails? All we need to do is consult, in cook-book fashion, JAN-STD-105 and we find that the O.C. Curve for the applicable sampling plan for these lots show that material 5% defective will be accepted 95% of the time (Figure III). But you say, "Suppose the Ace Company presents 8% material; what then"? Well, looking at the O.C. Curve

we find that lots of that quality will be accepted about 40 percent of the time. But is the Air Force willing to accept lots that bad and if not what will the AF do? The answer is that when we determine that quality is significantly worse than the AQL we will put in a more stringent inspection plan by which material 8% defective will be accepted only about 5% of the time and 5% material will be accepted about 55% of the time. (Figure IV). The conditions under which such "tightened inspection" is applied is well defined in JAN-STD-105. Likewise, under certain conditions - when the contractor produces particularly good material - reduced inspection may be used. (Figure V). The attached operating characteristic curves, if carefully studied, indicate the various features of normal, reduced, and tightened inspection.

Quality Standards

Now I know that some of you are asking how we "define clearly" the quality requirements of the Air Force for inspection by attributes. The question really is: "How do we implement JAN-STD-105?" The answer is that the Air Force establishes for each inspection item applicable Acceptable Quality Levels and also specifies the quality characteristics (Classification of Defects) for which inspection is required. Material which (for a class of characteristics) is not worse than an established Acceptable Quality Level will be accepted 95% of the time. The quality characteristics are grouped into different categories of seriousness by engineers who base their judgments on arbitrary criteria fixed by the Air Force as follows:

DEFINITIONS OF DEFECT CLASSES

Critical Defects

A critical defect is one that judgment and experience indicate could result in hazardous or unsafe conditions for

personnel using or maintaining the product; or, for major-end products, such as ships, aircraft, or tanks, a defect that could prevent the unit of product from performing its tactical function.

Major Defects

A major defect is a defect, other than critical, that could result in failure or materially reduce the usability of the unit of product.

Minor A

A Minor A defect is one that reduces the usability of the unit of product slightly.

Minor B

A Minor B defect is one that does not reduce the usability of the unit of product but nevertheless is a departure from specifications or drawings.

Consumer - Producer Relations

So far, then, we have outlined a plan which establishes clearly the relationship between the Government and the vendor insofar as quality requirements are concerned. As far as possible we have defined our quality requirements in quantitative terms because we feel that both the consumer and the producer are entitled to know the inspection hazards to which they are exposed. Producers require assurances that acceptance will not depend on the whims and fancies of the consumer; that all competitors, whether in Boston or Los Angeles, will be treated alike; and that the costs of quality evaluation will have some rational relationship to the importance of the item and to its cost of fabrication. Likewise the consumer should not be at the mercy of the producer. Enough of that, but before we return to the

Ace Company's nails it may be well to ask the jackpot question, "What is a fair AQL?".

Determination of AQL

Let us answer this question within the frame work of present Air Force procurement policy, i.e., there is no tie in between cost and AQL. In other words the contractor is responsible for delivering 100% acceptable material but the Government for practical reasons is willing to accept less than 100% quality. To the technical man it would appear that this problem is no problem at all. All you have to do, says he, is ask the aircraft nail designer just how much defectiveness can be tolerated and set the AQL there. Unfortunately we run into a particularly perverse breed of cat when we confront the aircraft nail engineer. All one gets is a lofty declaration that aircraft nails go into airplanes, airplanes fly, people get killed when they suddenly stop flying; therefore all aircraft nails must be perfect. It avails nothing to say that nails have been averaging 10% defective for years or that "perfection" of any sort is economically, if not physically, impossible. Next you might say, "Well, what can the industry economically produce? We will set the AQL there." The logic of this approach appears sound but it too has its weak points. First, reliable information regarding the quality capability of an industry is not usually available and is quite expensive to obtain. Secondly, it is difficult to say what the "average" quality of an industry really is without compromising the interests of manufacturers who have records of particularly high quality production. In the case of the aircraft nail we conducted a process average survey to determine exactly what the nail industry is now doing quality-wise. This survey was simply a study to determine the proportion of defectives in samples of nails taken from representative lots from each

of the cooperating manufacturers. Without going into the methods of conducting this survey here is what we learned at the Jack, Queen, King Ace and Deuce Aircraft Nail Companies:

PROCESS AVERAGE SURVEY
for
AIRCRAFT NAILS

QUALITY (PERCENT DEFECTIVE) OF MATERIAL
PRESENTED FOR ACCEPTANCE

COMPANY	MAJOR		MINOR-A		MINOR-B	
	A	B	A	B	A	B
JACK	7.8	8	0.6	5	2.9	3
QUEEN	25.9	8	25.7	5	3.7	3
KING	39.3	7	1.1	5	0.0	3
ACE	6.1	8	0.3	5	0.1	3
DEUCE	24.0	4	11.3	4	-	0

NOTE: Column A is percent defective.

Column B is number of characteristics inspected.

FIGURE VI

This survey was based on inspection for those characteristics listed on the C/D. However to present added inspection costs being charged to the Government, contractors were requested to report the incidence of defectiveness only for those characteristics which were, by standard practice, subject to Government and contractor inspection. (The AQL's finally proposed for the nail industry were 5% for Major characteristics and 6% for Minor A. A reasonable interim period is allowed for the contractor to make necessary adjustments to meet these AQL's).

The AMC Inspection Plan

It is well to pause now and see what kind of a plan we are offering the Ace Company. Here it is in outline:

THE AMC PLAN

Item -	Aircraft Nails
Lot Sizes -	Optional, provided material is homogeneous to the best knowledge of the contractor and Government.
Quality Characteristics Requiring Inspection -	Classification of Defects
Physical Tests & Associated Sampling Procedures -	(Omitted from this paper)
Sampling Plan -	JAN-STD-105 (For sampling inspection by attributes)
AQL's -	(Applicable to classes of characteristics) 5% Majors 6% Minor A — Minor B (No critical defects listed)

NOTE: Usually the details of an inspection plan are not included in specifications. Instead JAN-STD-105 is referenced and a standard inspection plan is prepared to implement the standard.

The Ace Company had little or no difficulty in meeting our AQL's; the Classification of Defects was found quite reasonable. There was only one real complaint against this plan; while it was protective, objective, and

standard, it was still too expensive. It cost more to inspect aircraft nails than to make them, a state of affairs that brings us to the subject of process control.

Process Control

This discussion has so far been restricted to problems of acceptance sampling which simply poses these questions: What is the quality of this material? Should I accept or reject it? There is, as you are well aware, another area of SQC called process control. This is by far the most fruitful field of Quality Control and in fact, to most production and industrial engineers, it is SQC. Unfortunately, statistical process control is not entirely under the scope of what the Air Force can consider its proper business. Without dwelling on this point at length, we can say that the Air Force is primarily a customer and, as such, should not concern itself with matters that are essentially in the field of production control, i.e., process control.

The Air Force Position on Process Controls

However, process control does present numerous problems because there is a tendency for manufacturers to consider process control in itself adequate protection for the Air Force against the acceptance of sub standard material. The Air Force, of course, wants to encourage the use of sound statistical process control methods but it would take a legion of engineers and statisticians to determine when a particular process control system actually gives the Air Force the protection it requires. To date we have steered a middle path by reducing acceptance sampling to a minimum when a process is known to be under effective controls. The term "to a minimum" means that we not only use "reduced" inspection as mentioned previously but on the basis of process control evidence we plan to group

lots into a "grand lot" and then sample this very sizeable population. In this way, of course, we cut our inspection requirements tremendously but at the same time we expose both the Air Force and the manufacturer to the hazard of rejection of an immense quantity of material. This problem becomes further complicated when the initial sub-lots of a grand lot are shipped before inspection of the grand lot is completed. Statistical methods, however, are available to minimize the hazards involved in a plan of this kind.

Results of an SQC Inspection Plan

SQC pays off when a manufacturer is truly producing under controlled conditions. A good example of how SQC thinking benefits both the Government and the contractor is the case of the ACE Aircraft Nail Company. Shortly after our original process average survey was made, this company established a complete statistical quality control organization. This new unit proceeded to institute statistical process controls practically on a blanket basis throughout the plant. I will leave it to you to judge the effectiveness of this program from the "before and after" results for nails outlined below:

AIRCRAFT NAILS	BEFORE PROCESS CONTROL			AFTER PROCESS CONTROL					
				FIRST CHECK			SECOND CHECK		
TYPE	MAJ.	MIN-A	MIN-B	MAJ.	MIN-A	MIN-B	MAJ.	MIN-A	MIN-B
	% Defective			% Defective			% Defective		
A	24.0	11.3	NO	0.4	0.1	0.4	0.8	0.0	0.8
B	4.5	2.5	INSPEC-	0.5	0.1	0.2	0.4	0.2	0.8
C	12.7	8.3	TION	1.3	0.0	0.7	NONE MANUFACTURED		

FIGURE VII

On the basis of this data, the Air Force authorized a "grand lot" evaluation plan for this company. The final result is of course that inspection costs have been cut significantly and production efficiency has reached a new high because operators anticipate "out of tolerance" trends in their machines. Quality Control thus achieves a positive effect. Evaluation is accurate, objective and economical; statistical controls are so sensitive that it is possible to adjust manufacturing processes before trouble develops. The further details of administration, cost adjustments, statistical methods, and revisions of specifications are so numerous that it seems best to leave it to you to choose those of greatest interest for discussion during the question period.

Conclusion - Resume

The function of SQC in Air Force Procurement is to assure the Air Force against the acceptance of sub-standard material. The methods by which such assurance is realized are based on numerous considerations of statistics, engineering and costs of inspection. The objective of SQC is to measure the quality of material by techniques which (1) make it possible for both the contractor and the Air Force to know their respective inspection risks,(2) identify specifically the quality characteristics which, for engineering reasons, are the foci of inspection, and (3) permits evaluation of quality at minimum cost. To the contractor this means (1) that in the future the measure of quality will be in more precise engineering and statistical language. (This, in all likelihood, will have an important bearing on competition). (2) Inspection, as a part of the overall SQC organization, will take on a positive function in manufacturing. (3) As far as the Air Force contracts are concerned, sound sampling inspection procedures will be written into specifications to

replace present nebulous inspection provisions. However statistical process control will not be considered to be of direct concern to the Air Force, but it may be an important factor in determining the severity of final acceptance inspection requirements.

As a concluding thought I should like to caution you that this review of the "Role of SQC in Air Force Procurement" makes SQC sound somewhat more "grown up" than it really is. To us, SQC is simply the common sense application of accepted statistical and engineering methods to the solution of problems of quality evaluation and control. However the job of revising specifications, preparing standard inspection plans, integrating process control and acceptance sampling, devising sampling plans for non-attribute inspection and tests, not to mention the job of "selling" SQC now and then, is so gigantic that we have only nibbled at the edge of the complete SQC program. But we can assure you that the Air Force is thoroughly quality conscious and is making an all out effort to use techniques for measuring quality which are in keeping with the scientific spirit of our times.

NO. 6
Price 25¢

MEASUREMENT AND CONTROL OF FLAVOR QUALITY

by

David R. Peryam
Chief, Acceptance Division
Quartermaster Food and Container Institute for the Armed Forces
Chicago, Illinois

for presentation at the

FOURTH NATIONAL CONVENTION
and
FIFTH MIDWEST CONFERENCE

of the

AMERICAN SOCIETY for Quality Control

June 1 and 2, 1950
Milwaukee Auditorium, Milwaukee, Wisconsin

FIFTH MIDWEST QUALITY CONTROL CONFERENCE
P. O. Box 1204, Milwaukee 1, Wisconsin

Measurement and Control of Flavor Quality

David R. Peryam
Quartermaster Food and Container Institute
Chicago, Illinois

Quality control can apply to almost anything that is manufactured, at least wherever it is possible to evaluate in terms of good and less good. Whether statistical quality control can apply so universally I do not know. Perhaps the common practices of S.Q.C., as we see it evidenced today, can not be transferred with efficiency to some industries or to certain critical aspects of their production. But it is becoming more and more evident that the basic idea of S.Q.C., that of practical use of scientifically sound measurement, has a very broad significance. Everyone may not find his problem fully solved therein, may not discover that the detailed methods he needs are fully matured and already made available by the pioneers of S.Q.C.; but he can find within its philosophy encouragement to tackle his own individual problem as well as demonstration of what the most valuable approaches are likely to be.

The field in which I work bears such a relationship to this new science. Through academic training, interest, plus the fortunes of employment I have always found myself concerned in production of food or related materials, and for the most part in fields related to quality. When I first began to hear about S.Q.C. I sensed an exciting parallel between the problems it handled with such aplomb and those with which I wrestled. I suppose my real feeling was, "Yes it's beautiful, scientific, and tempting, but how can that mathematical stuff do us any good here?" Yet upon my continued exposure to its ideas and its works I came

to appreciate the fact that I was already in the field without knowing it. That I am asked to participate here at the clinic sessions indicates to me that the professional S.Q.C. men concur in my belief.

Now for the moment I shall seem to digress. You may think I have become a bit elementary and obvious, but I want to make a point. When we talk about "quality" for quality control purposes, in spite of the many interpretations that might be given the concept, we know pretty well what is meant. Most interpretations of quality resolve into a matter of utility. If a product appears completely suitable for the purpose for which it is intended, by labelling it "good quality" we predict that it will give satisfaction to the ultimate consumer. Often the Quality Control man is lucky and can have very definite, if perhaps arbitrary, standards for what good quality shall be. Again, quality control may be required to discover within the production process itself characteristics in the product which are critical for this prediction of utility and, further, may have to devise ways to measure them reliably and efficiently. In all this the availability of methods of measurement is of vast importance. Consequently S.Q.C. has found most ready acceptance and its most precise application in those industries and for those processes where strictly objective measurement is possible.

Now consider production of food, beverages, and like products. Here we will find many aspects of quality that are accessible to objective measurement and therefore should also be subject to quality control by use of standard techniques. In this category will fall such factors as bacteriological and chemical purity, adherence to recipes and physical

specifications, and conformity to arbitrary physical grading standards. Also, wherever you have mass production it is usually possible to establish interim controls on the processes as such, making the reasonable assumption that such controls will somehow carry through in the quality of the product. This, of course, applies to food also.

But there is a facet of the quality of foods and like products where measurement is difficult. Moreover, this aspect is frequently, if not always, in the commanding position when it comes to determining total quality. This is the factor which may be called, simply, flavor. If a food is to be good the effect of its flavor must be positive, i.e., pleasant, or at least it must not have a negative effect. Yet as soon as we say "flavor" and admit that it is something more than a combination of chemical substances that impart smell and taste, we are in a new realm. Flavor is a concept that has meaning only in relation to the human senses for it arises in the interaction of the senses with the chemical environment. Therefore, its measurement must be in the same terms. Many factors which are subject to control in process by way of measurements which are strictly objective and precise contribute to flavor and their control represents partial flavor control. But a final evaluation of flavor must be in terms of human behavior.

Here is where quality control runs aground. Food producers can control processes, they can control properties related to flavor, but when it comes to the crucial aspect of the flavor itself, they are at a loss. Not that it is hard to know and evaluate flavor. Everyone can

and does do this but only on a personal basis. There is much variability in both perception and appreciation of flavor and every man has a tendency to feel that he is a judge unto himself. And so he is - for himself alone. But the manufacturer can not afford to define quality in terms of the responses of a few persons. Evaluating flavor so as to take account of the probable reaction of the consumer _en masse_ can not be done in this easy, naive way. To predict mass behavior, we must sample behavior, we must measure it by rigorous, standard procedures, and we must be able to understand the significance of what we have done.

Among food processors and research workers attitudes toward the problem of flavor control vary from the supremely confident to the apathetic. The former goes with the belief that individual judgment is entirely sufficient, so we do not need a careful approach; the latter with the belief that the problem is not capable of solution, so we should not waste time on it. That both attitudes have arisen is understandable in light of the present situation. Which situation can best be described in a flat statement: _A standard methodology for dealing with flavor problems does not exist._

However, such methodology is possible. It is even now being developed by some middle-of-the-roaders; yet it is still poorly formulated. Also, knowledge of even what little has been done is not widespread. Until it is flavor control will not be generally scientific and effective.

Tonight I propose to tell you something of what is being done in this field. First, I will present in some detail what is undoubtedly

the best example of scientific flavor control in industry today. Then I will discuss some experimental work that is suggestive of the way further progress is to be made.

That best example of flavor control is the system used by the Seagram-Calvert Distilleries. Their production organization changed over to a scientific approach to flavor problems about eleven years ago and has used the present system about eight years. I was with their quality Laboratory for four years during the period when the method was being developed and installed, and came to know it first-hand. Since I left the organization and have had an opportunity to learn more of flavor control problems and of their solutions among food producers at large, I have come to appreciate the effectiveness of the Seagram-Calvert methods.

At each of the company's operating plants there has been established an organization which is designated The Quality Laboratory. These are essentially psychology laboratories rather than whiskey laboratories; their basic procedures are determined by that science and their function is to use human observers in development and process control work and, finally, to handle the important task of product acceptance. It is the method used for the latter purpose which I will describe here.

The traditional method of quality control in the liquor industry is the expert system, in which one or a small group of expert tasters attempt to handle the job by means of particularized knowledge of whiskey and by individual sensory skill. This is a perfect example of

the overconfident approach. The experts must presume both unerring judgment and unerring sensory perception. The laboratory methods, in contrast, put taste-testing on a level where both validity and reliability of results are markedly improved.

Control testing implies a standard. The subjective standard used by the consumer, and by the expert taster, consisting as it does simply of a knowledge of what is good, is too vague, too subject to chance variation, to be an accurate guide. The solution was to create a physical standard embodying the desired flavor. Then testing can be done in terms of difference from this standard. Capricious change is thereby avoided and one also takes advantage of the fact that difference testing can be much more precise than the testing of absolute values. Originally to obtain the standard for a given product the best combination of ingredients possible, considering time and whiskey stocks available, is worked out by a series of taste preference tests using observer groups representative of the customer public. This standard is maintained by an extension of the same procedure used for acceptance work and any variation therefrom is made by prior decision.

In the production process the whiskey is blended in lots of from 8 to 10 thousand gallons according to a specific formula worked out by the Quality Laboratory which expresses the master standard in terms of the particular whiskeys available. After processing is completed a sample of each tank lot is submitted to the laboratory for an answer to the question, "Can this lot be distinguished from its standard by sensory test?" If the answer is negative, the lot is accepted; if the answer is

6

affirmative, the lot is rejected and must be reworked. When a rejection occurs, it indicates that either one or more of the whiskeys or other raw materials used were non-standard or else that there has been an irregularity in the production process. In either case the causes can be sought and, if identified, can be eliminated.

This system provides for a complete inspection of outgoing product at the completion of the blending process, since no lot of whiskey is sent to the bottling tanks until sampled and approved. The possibility of flavor contamination during the bottling operation is but slight. However, the laboratory also makes periodic checks of samples taken from the bottling line.

With this general background established we will go on to specific description of the test used, including some discussion of controls which are necessary.

Without doubt, the observers, i.e., the persons who do the testing and judging, are the most important aspect of the laboratory. Without these individuals, without their sensory and perceptual abilities manifested under conditions of proper control, the system is useless. Their selection and training cannot be haphazard and their performance must be continuously subject to review. Ten observers comprise the normal testing group. They are usually women, originally hired to work on the bottling lines, who have been assigned to the laboratory on a trial basis and have proven their ability. Very likely an observer's sense of taste will be more acute than average, but the selection is made upon the basis of demonstrated ability to detect differences between whiskies.

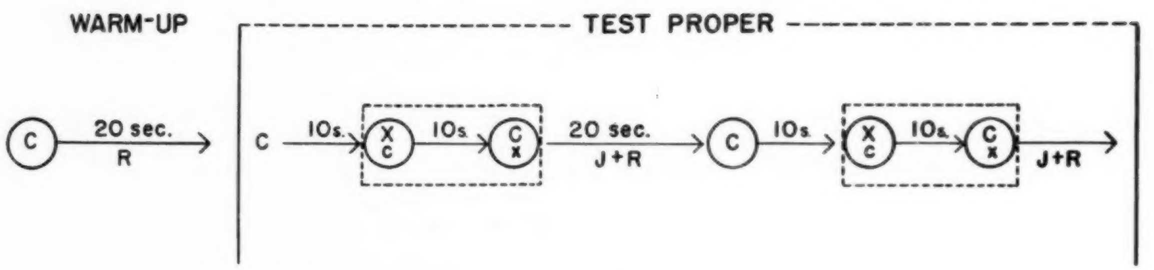
It should be emphasized that these people are not considered "whiskey experts" since their usefulness does not depend upon any kind of unusual knowledge or skill. Not only do they know what they are doing, but we know also, and a method is readily available for training replacement.

At present a taste test only is used in acceptance work. At various times an odor test has been used in parallel. However, the two correlated highly and, since the odor results were more variable, the latter test was dropped in the interests of inspection economy.

Basically, the test is quite simple. The observers taste 4 cc samples of whiskey from ordinary shot glasses. These samples are presented in a controlled order and the observer is required to make certain statements about them. Special apparatus has been designed to make administration of the test speedy yet accurate and the observers have been trained to follow procedures in testing which have proven to give best results.

Figure 1. diagrams the procedure of a single test. It shows the series of seven taste samples which is given to an observer at a sitting. This procedure has been given a name - the "duo-trio" test.

Prior to the start of the test proper the observer gets as a "warm-up" sample of 4 cc's of whiskey "C," which is the standard. The "warm-up" has this function: The first sample which is tasted may seem quite different from the others merely because of its temporal position and because of contrast with the previous neutral condition in the mouth. The "warm-up" readies the mouth, making conditions more nearly constant for the rest of the samples. It is followed by a minimum interval of 20 seconds



C-sample of standard whiskey X-sample of unknown whiskey
R-indicates rinsing with water J-indicates judgment on unknown pair

FIGURE 1. DIAGRAM OF THE SAMPLE SERIES WHICH CONSTITUTES THE 'TEST' WHICH IS TAKEN BY AN OBSERVER AT ONE SITTING.

during which the observer rinses with distilled water which is always available in quantity in the testing booth. In the test proper "C" is given first, followed, at 10-second intervals, by the two members of the "X-C" pair given in an unknown order. Next, there is another 20-second interval during which the observer gives his judgment and again rinses. Then the process is repeated, i.e., another "C" followed by a second unknown pair for judgment. The time required for the entire test is about two minutes. Test series for a given observer are separated by a minimum interval of 15 minutes to give his sensory mechanisms time for complete recovery.

The crux of the whole matter is the response. In this test the observer is fully aware of the entire series, i.e., he knows that the first sample is a "warm-up" and that in the test proper the first sample of a trio is always "C" and is followed by a pair in unknown order. He knows that, within the pair, one member will be identical with "C" but that the other may differ from it. He must state which of the two is the "different" sample. "But," you may object, "suppose he thinks they are identical and are both like the standard?" This does arise frequently, for "X" will be very close to "C" if the lot from which it was drawn has been properly blended. However, the test does not permit a "same" judgment; the observer must give his best guess. When we turn to analysis of the data we will see why this is necessary. It can be shown that this forcing of a decision is not unrealistic. Rather, it makes for finer discriminations. Not infrequently the test as a whole proves that a difference exists even though no single observer has been certain of

his judgment.

The data are recorded simply as correct or incorrect for each judgment according to whether the observer has properly identified the "X" or not. The complete control test on a given production lot consists of ten of these individual test series using ten different observers. Since each person gives two judgments there are 20 items for final analysis. When results with the 20 judgment test are doubtful, it may be extended to 30 judgments by having five observers repeat.

Controls which are vital if the test is to be effective fall into two general classifications. First there are those relating to the test samples, "C" and "X," themselves. They must be identical in every aspect under laboratory control to assure that the observers will not get irrelevant cues.

1. Temperature must be equal. It has been found that normal room temperature is optimum for difference testing.
2. Alcohol concentration (proof) must be identical since a difference of about 2% of proof is detectable by a trained observer.
3. There must be no variation in color.
4. All samples must contain the same volume.

The second class of controls is larger. Their need is dictated by common sense, refined by the science of psychology, to assure that the observer may work under optimum conditions. Most of them relate to the factors of adaptation and fatigue.

1. As previously explained the "warm-up" sample is given to make adaptation uniform.

2. The water rinses are employed to slow down the rate of adaptation to the flavor of whiskey, because, as it proceeds, finer differences disappear.

3. Also to retard adaptation, there is a "no-swallowing" rule. The entire sample is expectorated. When the whiskey reaches the deeper regions of the buccal cavity and the throat it can be perceived more fully. However, in that case rinsing is relatively ineffective and the net result is reduced sensitivity.

4. The samples are reduced to 45° proof, which is half the usual whiskey strength. At full strength, whiskey has such a strong flavor and tactual quality as to interfere markedly with taste perception.

5. Time intervals between samples must be controlled with two conflicting objectives in mind. At least a minimum of recovery time must be allowed but intervals cannot be extended to the point where forgetting becomes important.

6. The order of presentation within the unknown pairs is determined by chance. If it is the same all the time or follows any set pattern, constant errors might arise.

7. The testing situation must be kept free from distractions, both sensory and ideational, to permit observers to operate at the optimum. For example, smoking is not permitted in the laboratory. Production personnel are not allowed to discuss results with the observers.

8. Motivation is important. To make the system work the observer must consider each unknown pair a challenge. It is made explicit that the criterion of job success is the ability to make a good percentage of "right" responses. At the end of each test the observer is told his results and a continuous record is kept of each observer's percentage correct. She knows that the penalty for falling singificantly below the performance level of the group is return to the bottling line. Also, a day by day competitive spirit is fostered among the observers.

How do we use these data for control? Remember that all of these judgments may be guesses or, on the other hand, they may represent a greater or lesser degree of certainty on the observers' part. However, all are treated as being of equal value. Also each judgment is treated as a separate datum even though it is a member of a pair and there is probably some degree of correlation.

If blending of the lot under consideration has been normal the "X" will be identical in taste with the standard "C," or quite close to it, and most of the judgments will be, actually, guesses. However, over a period of time lots will appear which cover quite a range of flavor variation. As difference from the standard increases, guessing will become less frequent and therefore the percentage of correct responses will be higher. How high must this percentage go before we can be certain that, considering the group of responses as a whole, they no longer could represent guess work? How high before we can safely say, "Here is a true difference?"

To answer this question we look at a hypothetical set of data. What can we expect to happen when the samples are identical? Here we have a situation analogous to the penny-tossing experiment where we record the percentage of heads in twenty tosses, for there are but two alternatives, and each observer has a 50-50 chance of being right on each judgment. Following this analogy further, we know that if we run our taste test on identical "X-C" pair and then repeat the test many times over, recording the percentage of correct responses obtained each time, these percentages-correct will tend to distribute on the normal curve with 50% correct as the mean. (It has been established by experiment that the results on identical samples will so distribute.) Now we inquire as to the frequency with which any given percentage-correct will occur.

We analyze our theoretical distribution in terms of the Standard Error, using the formula:

$$\sigma_p = \sqrt{\frac{p \cdot q}{N}}$$

σ_p = Standard Error of a percentage
p = percentage-correct
q = (1-p) or percentage incorrect
N = number of cases

Here the true mean is 50%. Therefore,

p = .50
q = .50
N = .20

$$\sigma_p = \sqrt{\frac{.50 \times .50}{20}} \quad .112 \text{ or } 11.2\%$$

The significance of the σp is this: Upon repetition of the test on the identical samples, the result will fall within certain σp limits from the mean percentage with a frequency which can be estimated by reference to the standard tables of area under the normal curve. Figure 2 is a diagram of the situation under discussion.

The mean percentage-correct is at 50%. Referring to the tables, 68.2% of all tests will fall within $1\sigma p$ each way from 50% -- or between 61.2% and 38.8% correct judgments -- and 95.4% will fall within $2\sigma p$, or between 72.4% and 27.6% correct judgments. The expectancy that the result of the test will fall beyond any given "σp" point is represented by the area under the curve beyond this point. For example, we could expect 2.3% of all tests to fall beyond 72.4% correct judgments and the same percent to fall beyond 27.6% at the lower end of the curve.

Table I shows these calculations for all possible percentages using twenty judgments, and, for the thirty judgment test, the figures near the region which is critical for purposes of this test. Only half of the symmetrical curve is tabulated. In these calculations we have assumed a normal curve for the data. Since we are dealing with a binomial expansion this assumption is not strictly correct but it is a close enough approximation for this particular application. Columns 1, 2, and 3 show, respectively, total judgments, number correct, and percent correct into the σp distance from 50%. Column 5 shows the percent of area under the curve between 50% and this point and Column 6, the percent beyond. Column 6 has another meaning since it also gives the chances in one-hundred of equalling or exceeding the given percentage correct.

TABLE I

RATINGS AND AREAS UNDER THE NORMAL CURVE
CORRESPONDING TO GIVEN PERCENTAGES CORRECT

Total Judgments	Number Correct	Percent Correct	SE Distance from 50% (6p rating)	Percent area between 50% and this point	Percent area beyond this point
20	10	50.0	0.0	0.00	50.0
20	11	55.0	0.4	15.6	34.4
20	12	60.0	0.9	31.6	18.4
20	13	65.0	1.3	40.3	9.7
20	14	70.0	1.8	46.4	3.6
20	15	75.0	2.2	48.6	1.4
20	16	80.0	2.7	49.7	0.5
20	17	85.0	3.1	49.9	0.1
20	18	90.0	3.6	50.0-	Very small
20	19	95.0	4.0	50.0-	" "
20	20	100.0	4.5	50.0-	" "
30	18	60.0	1.1	36.4	13.6
30	19	63.3	1.5	43.3	6.7
30	20	66.7	1.8	46.4	3.6
30	21	70.0	2.2	48.6	1.4
30	22	73.3	2.6	49.6	0.4
30	23	76.7	2.9	49.8	0.2
30	24	80.0	3.3	50.0-	Very small

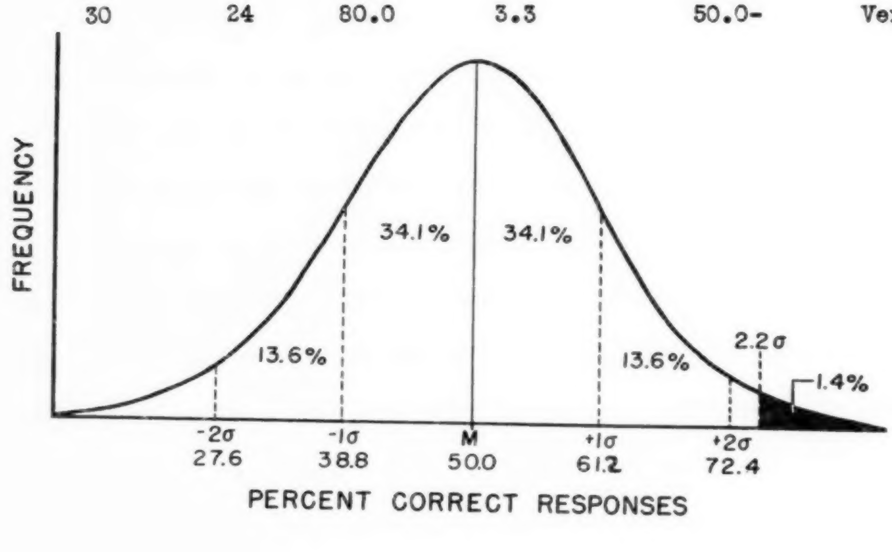


FIGURE 2. THEORETICAL DISTRIBUTION OF PERCENTAGES CORRECT UPON REPETITION OF THE TASTE-DIFFERENCE TEST ON IDENTICAL SAMPLES USING 20 JUDGES. SHADED AREA REPRESENTS PRODUCTION RISK WITH THE 2.2σ REJECTION POINT.

To test for difference we compare any given result on an unknown lot with this theoretical distribution and find out how often the result might arise with identical samples. If the latter frequency is low we conclude that there is a true difference.

The reader may be interested in the significance of percentage-correct results below 50%. They usually represent nothing more than chance expectancy with identical samples. However, if any percentage is lower than chance reasonably would allow, we must conclude that the test was faulty -- either that the observers were making some type of constant error or else that for some reason they were trying to defeat the test in order to approve the sample. In either case the test is repeated with additional precautions.

With this analysis behind us we still have the problem of selecting a workable rejection point, one which gives assurance of customer satisfaction, yet one which can be met by production without too great expense. The breaking point at present in use is $2.2\sigma p$, which point is indicated in Figure 2. Any result below this point approves the lot, at or beyond, rejects it. Referring to Table I we see that, at this point, there are only 1.4 chances in 100 that the lot could be identical with the standard. Note, however, that this does not explicitly tell anything about the degree of difference between the two. It is a logical inference that greater "σp" ratings reflect larger degrees of difference, but the method does not require this to be true. In effect we have said only that it is highly improbable that the lot is identical with the standard; definitely we have not said that the difference is such that we should always expect

such a result. In fact a repetition of the test is quite likely to result in a lower percentage, although never as low as 50%.

Following are examples of the system in operation. They can be checked by reference to Table I.

Example 1:

Number of judgments	--	20.
Number correct	--	13
σp rating	--	1.3
Disposition	--	The lot is approved.

Example 2:

Number of judgments	--	20.
Number correct	--	16
σp rating	--	2.7
Disposition	--	The lot is rejected and must be reblended.

Example 3: Let us say that the 20-judgment test showed 70% correct. The σp rating would be 1.8. This is considered doubtful and the test is continued.

New total	--	30.
Number correct	--	19
σp rating	--	1.5
Disposition	--	Approved

Example 4: The 20-judgment test shows 75% correct -- σp rating, 2.2. Again this is borderline and the test is extended.

New total	--	30.
Number correct	--	21.
σp rating	--	2.2
Disposition	--	Rejected and reblended.

This method does not attempt to "prove" that the production lots are the same as the standard. Rather, it implies identity for practical purposes, through failure to prove that a difference exists. This is the customer's assurance of quality. With an approved rating of 1.8σp, it is still quite possible that there is a true difference, but since, even

under the optimum laboratory conditions, we cannot prove the fact, we know that under field conditions of liquor consumption, the difference will be unnoticeable.

Replenishment of standards represents a more difficult problem than the testing of production lots because here the accumulation of even slight differences, over a period of time, could lead us away from the original taste quality. The solution is to extend testing of potential standard to five times the usual test-length, 100 judgments, and to accept lots only where results are internally consistent and where the combined percent-correct lies close to 50%. Then we do have fair statistical assurance of identity.

Note that production must take the risk of having rejected, on the average of 1.4 times in 100, lots which are identical with the standard. There is a much greater, but unknown, risk of rejecting lots which differ in such slight degree that they would have no adverse effect on sales. Of course, in the final analysis limits are set, not on the basis of statistical theory alone, but also on the basis of what production can achieve, and at such a point that the cost of reblending, and the incident delay, do not become prohibitive. At the outset of the program limits were much wider than at present, but, under the strong motivation provided by the system, and aided by the definite information it could give, the production department did such a good job of process control that limits could be progressively narrowed.

We should note that a producer of alcoholic beverages can use this system to greater advantage than most food processors because his product

is quite stable. If you put a sample of whiskey in a sealed pyrex bottle it will retain its full flavor indefinitely. There are no worries about deterioration or possible contamination. This permits the distiller to maintain an actual physical standard over a long period of time. Few food products are like this, time is an important factor in their flavor quality. More often than not there is no assurance that a physical standard will remain constant in flavor for an extended length of time. In such cases the difference testing procedure exactly as used by Seagram's is hardly applicable and we must seek stable reference points elsewhere.

The run-of-the-mill solution has been to use a completely subjective standard as represented by the memories of a few trained individuals. In other words, you rely on persons who simply remember the flavor. However, it has been demonstrated many times that this is a make-shift arrangement. Its measurements are unreliable and the standard it represents is not standard enough to be fully effective.

Some methodological work is now underway at the Quartermaster Food and Container Institute that has a bearing on this problem. I will tell you briefly of two phases of our work. However, I wish to emphasize that they do not represent finished flavor control systems but are only suggestive of the directions which such work may take.

The first relates to dried milk, but might be applicable to any product where there is the same difficulty of maintaining a standard. Dried whole milk is one of the problem children of the dairy industry. It shows high variability in quality, ranging from material fit only for animal feed to that which might even complete for the fresh milk market. A good

6

deal of effort has gone into flavor evaluation in attempts to improve and control quality, most of it being done with the subjective standard and expert panels. This work has assisted process research in slowly raising standards, but it has not given effective control of flavor quality.

Our experiment started with the reasonable assumption (which has been verified by consumer preference tests) that quality in a dried whole milk depends upon how closely its flavor approximates that of fresh milk. Even though this is generally conceded, such knowledge seems to do little good in a test involving subjective evaluation, probably because the sensory difference between the two types of milk is so marked.

We set up an index of quality called Dilution Number. This index we defined as: "The percent of dried whole milk in a mixture of this substance with a fresh milk standard such that the difference in flavor between the mixture and the standard is barely detectable." In other words, how much of the dried milk must be put in fresh milk before the flavor change is detectable.

To establish the index for a given sample, we set up a series of dilutions, differing by small increments of percentage of dried milk, and such that their flavor seemed to be close to the point of a being just detectably different from fresh milk. These dilutions were tested one after another for difference against the fresh milk standard using the duo-trio test with a total of 20 judgments. The critical dilution was that which gave just 15 correct responses out of 20. Refer to Table I and you will note that this is the first point which goes beyond 2σp.

A typical set of results is shown in Figure 3, which also shows calculation of the Dilution Number when no dilution gave exactly 15 correct responses.

Our experimental work with this system was encouraging, though far from adequate for "proving" it. Table II summarizes the results from one experimental run. We established DN's for 11 samples of dried milk selected so as to represent a range of quality, re-established them later to check reproducibility, and also submitted them to that ultimate board of appeal, a consumer preference test. The reproducibility of the index was found to be high, as evidenced by a correlation of +.96 between the first and second series. There was a correlation of +.89 between the rank orders of the samples by Dilution Number and by consumer preference ratings, which was equally encouraging.

A logical next step in verification of this method would be to have a manufacturer use it routinely in order to find out how effective it would be in a practical situation. Of course, a control laboratory would not be able to do as much work on each sample as we did. But here is the way the system might be used: By careful test the laboratory would establish a DN for the flavor quality the plant could consistently produce. Let us say the DN turns out to be 30. They also would establish the DN for quality below which they know they would not want to go. This might be 25. The following tests would be run on a representative sample of each lot, or, for a continuous process, at fixed time intervals:

6

TABLE II - DILUTION NUMBER AND CONSUMER ACCEPTANCE
RATINGS OF ELEVEN DRIED MILKS

Sample Code	First DN* (a)	Second DN (b)	Average DN (c)	Consumer Acceptance** (d)
A	30.0	32.5	31.2	4.9
B	27.5	25.0	26.2	4.6
C	25.0	22.5	23.8	5.2
D	22.0	25.0	23.5	5.1
E	23.0	22.5	22.8	4.7
F	23.0	20.0	21.5	4.5
G	19.0	23.0	21.0	3.6
H	13.0	15.0	14.0	4.4
I	10.0	14.0	12.2	3.2
J	8.5	8.0	8.2	3.0
K	0.3	0.5	0.4	2.4

Product moment correlation between (a) and (b) = +.96
Rank order correlation between (c) and (d) = +.89

* - Dilution Number
** - By a 9-point "like-dislike" scale

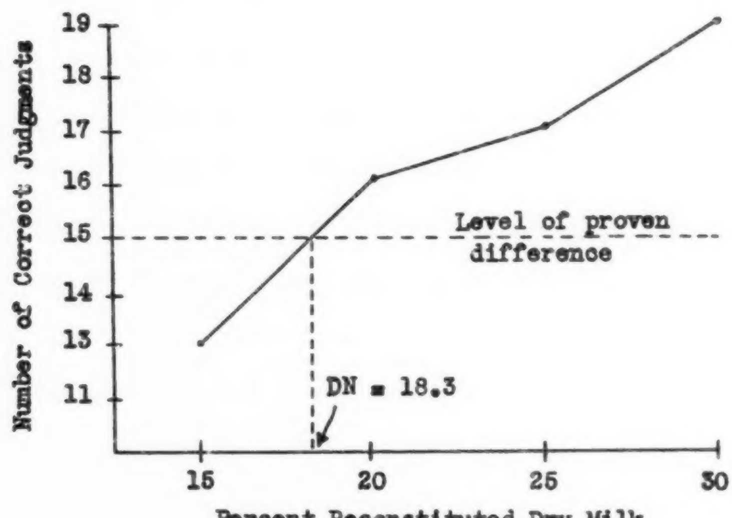


Figure 3 Method of determining Dilution Number

Test #1. At DN = 30 (Fresh milk <u>vs.</u> a mixture of 30% product plus 70% fresh milk.)

a. Result below $2\sigma p$ - approval. Production is in control; no more testing necessary.

b. Result $2\sigma p$ or above - non-approval. Test further.

Test #2. At DN = 25

a. Result below $2\sigma p$ - approval. However, the product is suspect and the process should be investigated.

b. Result at or beyond $2\sigma p$ - rejection. Product is sub-standard and process is out of control.

The second problem which we worked on at the Quartermaster Food and Container Institute is that of establishing a valid measure for acceptance of a product for which no natural standard exists. Soluble coffee is a good example, bouillon another. We are working on both of these products at the present time. The use of an arbitrary physical standard in such cases might result in great economic waste by rejecting material which, though quite different in flavor from the standard, would be equally as good in terms of the mass consumer response. In this case the testing philosophy moves closer to the final criterion of preference. A consumer's reaction will be one of liking (or appreciation of good quality), which will lead to acceptance, or else one of dislike (or awareness of poor quality), which will lead toward relative non-acceptance. Here, to accomplish the necessary task of making a direct estimate of the consumer reaction, a method is used in which precision is much poorer than in the difference testing previously discussed, but where validity can be easily upheld. It utilizes the quality judgments of a special panel of

observers as expressed on a rating scale.

We use a panel of 20 persons who have been selected on the basis of ability to make fine discriminations of differences in the product under investigation (which is the same basis of selection as used for the Seagram-Calvert observers) and also on the basis of ability to reproduce judgments of quality. These people are familiarized with the product by taste-testing a variety of samples covering the normally expected range of quality.

Figure 4 shows the test form which is used in evaluation of soluble coffee, including both instructions and the rating scale. Note that this scale calls for a judgment of quality rather than for a "like-dislike" response such as would be used in a consumer preference test. Panel members are instructed to consider good quality that which they believe will have high consumer preference even though this may not agree with their personal preferences in all cases. They work independently while testing. Values assigned by the 20 people are then averaged to get the final rating.

The panel will generate its own standard for acceptance by evaluating representative samples of a number of brands, or grades, of a product. As many samples are tested as necessary to give a reasonably good estimate of the spread of mean scale values which may be expected considering everything on the market, and sufficient data are taken to permit a good estimate of the precision of a single mean value. Table III shows a set of typical data from this method and how they would be used to set an acceptance standard.

TABLE III - EVALUATION OF A SERIES OF SAMPLES OF SOLUBLE COFFEE ON THE 9-POINT JUDGMENTAL SCALE BY A 20 MEMBER PANEL

	Mean Rating				Standard Deviation			
Sample	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	Sample $\bar{X}$	σ_1	σ_2	σ_3	Sample σ
A	6.9	7.3	7.0	7.1	.9	.8	1.0	.9
B	7.0	6.8	6.7	6.8	1.2	.9	.9	1.0
C	5.9	6.3	6.2	6.1	.7	1.0	.9	.9
D	5.0	4.9	5.2	5.0	1.0	1.2	1.2	1.1
E	4.7	4.3	4.8	4.6	1.3	1.1	1.2	1.2
F	3.8	4.2	4.1	4.0	1.1	1.4	1.0	1.2

Grand mean of all samples - 5.6

Average of sample σ's - 1.05

σ_M - .23

$3\sigma_M$ - .69

COFFEE RATING SCALE

You will receive one or more samples of coffee, one at a time, which you are to judge as to quality. Evaluate each by the phrases printed below the scale. Write the sample code at the left and indicate your judgment by a check mark on the scale.

You may swallow the coffee if you choose but treat all samples alike. Rinse your mouth thoroughly after each sample.

This is not a preference test. You are to judge each coffee in terms of how acceptable you believe it would be to most consumers.

Sample Number	1	2	3	4	5	6	7	8	9
___	Extremely Poor	Very Poor	Poor	Below Fair Above Poor	Fair	Below Good Above Fair	Good	Very Good	Excellent
___	1	2	3	4	5	6	7	8	9
___	1	2	3	4	5	6	7	8	9

FIGURE 4 - RATING SCALE USED IN JUDGMENTAL EVALUATION OF SOLUBLE COFFEE

The panel tested each sample of coffee three different times. The mean scale value and standard deviation are shown for each test, followed by a "Sample $\bar{X}$" and "Sample σ" which are in each case the averages of the three values. Sample σ's for the six samples are averaged, giving a value of 1.05, which is an estimate of the variability which can normally can be expected in the distribution of judgments when this panel rates a coffee. From this we calculate the standard error of the mean when 20 judgments are taken:

$$\sigma_{M(20)} = \sqrt{\frac{1.05}{20}} = .23$$

This is the estimate of the normally expected variability of a mean rating.

The "Sample $\bar{X}$" values serve as a guide in setting the quality standard although just what level should be selected still remains a matter for decision. The "grand mean of all samples" is one logical value to use. This would represent a decision to accept anything which was not provably in the low quality range. This plan would have particular merit if the products which were tested to establish the standard gave mean scores which were distributed normally. Top quality could be demanded by setting the standard at 7.1, which was the best mean score found with the coffees tested.

Still another method that has good justification is this. Note that two samples in the series tested here have mean ratings that lie more than $3\sigma_M$ below the "grand mean of all samples." They are obviously of inferior quality and might for that reason be disregarded in setting the standard

6

value. For example, here one would recalculate the mean for the best four samples and get 6.25 as the standard score.

In any case, after the standard score is set, the point of rejection on the basis of a single mean rating by the 20 member panel would be set at 0.69 points ($3\sigma_M$) below the standard score. Perhaps it could be set at $2.5\sigma_M$ below the standard if one wanted to tighten up on control.

This system is now being put to practical use. Many military specifications for food items carry the general requirement that the material must be of "acceptable flavor." This is above and beyond any of the physical or chemical specifications that are subject to the usual tests. Our organization has the responsibility of determining, for beverage items, whether pre-bid samples are acceptable in flavor. Here is exactly the situation which I have described on a theoretical basis. There can be no single physical standard, since items may vary considerably in flavor yet be equally acceptable. Again, we would not be justified in setting a standard any higher than the "grand mean of all samples" considering the usual wording of such specifications. If we reject any sample we must be able to prove that it is definitely below standard, or, as we would have to say, "Non-acceptable."

We can not say yet how practical the method will be as we are still in the process of selecting panels and setting up standards, but work is underway on synthetic beverage bases, soluble coffee, and bouillon. Once we have selected our panels and have calibrated them on the particular products, then we must wait and see how effective a control will result. One thing we know, however, is that for all its possibilities of inaccuracy,

this system will be far better than the semi-arbitrary "expert" type of sample screening it supplants. Such a rational, explainable method not only will give us some measure of assurance in our decisions, it will also help the manufacturer for he will be able to find out what the word "acceptable" in the specification means.

My purpose in this discussion has been two-fold. First, to tell you something of the difficult problem of control of flavor quality; and, second, to indicate that at least some people are approaching that problem with the same scientific rigor that is doctrine in the field of Statistical Quality Control. I will readily concede that it is unlikely that flavor control will ever attain the precision that is possible where control characteristics can be measured more objectively. However, great improvement can be made over the methods that are common practice in industry today.

NO. 7
Price 25¢

BALANCING AND RANDOMIZING IN EXPERIMENTS

by

Warren R. Purcell
Manager of Quality Control
Lighting Division
Sylvania Electric Products, Inc.
Salem, Massachusetts

for presentation at the

FOURTH NATIONAL CONVENTION
and
FIFTH MIDWEST CONFERENCE

of the

AMERICAN SOCIETY for Quality Control

June 1 and 2, 1950
Milwaukee Auditorium, Milwaukee, Wisconsin

FIFTH MIDWEST QUALITY CONTROL CONFERENCE
P. O. Box 1204, Milwaukee 1, Wisconsin

BALANCING AND RANDOMIZING IN EXPERIMENTS

By

Warren R. Purcell

Manager of Quality Control, Lighting Division

Sylvania Electric Products Inc.

Fellow, American Society for Quality Control

Boston Society for Quality Control

Introduction

The Modern Quality Control Engineer is a vital cog in the wheels of today's industry. His services have become all-important in the ever-present problem of identifying, eliminating, and preventing the causes of excessive variability in manufacturing. By the judicious and skillful use of various statistical and graphical tools, such as histograms and control charts, he has been able to help substantially in identifying causes which elude identification by old-fashioned methods.

Occasionally, however, he has encountered difficult problems of cause identification, especially in the more complex and technical processes. The solutions of such problems often require full-scale engineering investigations. In the early days of Modern Quality Control, such investigations were considered a separate problem to be solved by Material and Process Engineering. It is now realized that the Quality Control Engineer has at his command additional statistical and graphical tools, such as variance analysis and correlation, to assist materially in the interpretation of the data yielded by engineering investigations.

A necessary requirement for the effective use of these specialized tools is a "way of thinking" in terms of variability and probability. Herein lies the real value of the Modern Quality Control Engineer. Because of his way of thinking, his services are valuable not only in the interpretation of data from experiments, but also in the planning and executing of the experiments themselves. Thus his usefulness is now extending beyond the problems of day-to-day process control, important as they are, to the fields of research, development, and design. Extraneous causes of variability not only disturb established processes, but also enter into experiments and often make the results tell costly lies. The fallibility of human beings and the perversity of inanimate objects are constantly rearing their heads to plague the investigator. The Quality Control Engineer is, in many cases, the one person in the organization fully trained to cope with these intruders.

Fallacious Procedures

One of the most troublesome of the perversities of inanimate objects is the tendency of true causes to disguise themselves and to try to appear as something else. Consider a few illustrations.

Not long ago a packing engineer devised a new package for a certain product. Adoption of the new package would have saved several thousands of dollars annually. Before final adoption a transportation test was run to indicate whether the proposed packing would properly protect the product. In the test 1000 inspected units were shipped on a round trip in the proposed packing and 1000 inspected units in the standard packing. Upon their return they were again inspected. In the standard packing one unit was found to be defective; in the proposed packing 10 units were defective. The decision to forego the contemplated saving had almost been reached when it was pointed out that the defects could have been due to weakness in the product rather

than to inadequate packing protection. Further investigation disclosed that this was actually the case, and that the proposed packing could, in fact, be adopted. The lie told by this experiment could have cost thousands of dollars.

At another time an engineer ran a test to determine the relative strengths of lamp filaments made from two different types of tungsten wire. To test the strength, 20 lamps were manufactured using one type of wire, and 20 using the other type. The lamps were then put through a standard "bump" test to determine the relative percentages of filament breakage which would result. In this test the glass bulb occasionally shatters and spoils the test of one of the lamps. But in this particular test, a strange result happened. Of the 20 lamps made with one type of wire, none of the bulbs shattered; of the 20 lamps made with the other type, 10 of the bulbs shattered. The conclusion indicated was that one type of wire caused brittle bulbs. The lie told by this experiment was so ridiculous that it fooled no one. There was, therefore, little lost in this case, unless we count the time, effort, and material expended, plus the loss of faith by the engineer in experimental results in general.

Another case is of a somewhat more familiar type. The life of a particular size of lamp was running shorter than its design called for. A suspect was the speed of the sealing machine. An experiment was therefore run in which lamps were manufactured both at the standard speed and at a slower speed. Life tests were completed, and the lamps made at the slower speed had substantially longer life. Even though it meant a loss in production, the anticipated increase in life was important, and the machine speed was therefore reduced. Unfortunately, the life remained just as it had been at standard speed; nothing was gained when the speed was reduced in actual production. The lie told by this experiment cost about ten thousand dollars in lost production, and the problem of short life remained unsolved.

These are actual cases from our files. Others could be cited. Such cases indicated to us the existence of serious fallacies in our procedures. If such obviously erroneous conclusions could be drawn, in how many cases were less obvious, but no less costly, wrong decisions being made? Every decision which is based on experimental evidence is exposed to this same hazard. What can be done about it?

Quantity of Evidence

One approach is to require that every important experimental decision be based on a large quantity of evidence. "Can you repeat the experiment (several times, perhaps) and keep getting the same result?" is the question often asked by one who uses this approach. Perhaps the suggestion is made to "try it in actual production and see what happens". But these procedures can be costly. If they produce positive evidence, they may be well worth while. But often the results become conflicting and inconclusive. A relationship that is established by today's experiment does not hold tomorrow. Engineer Smith finds that an increase in furnace temperature at a certain process stage will yield a more wear-resistant surface, but Engineer Jones finds that such a temperature increase is without effect on surface condition, or even makes it less wear-resistant. Such conflicting evidence is caused by the same kind of faulty procedures that make the individual experiment tell such costly lies. Causes not only try to disguise themselves, but they sometimes hide completely, and at other times come into full view, all with a bewildering sequence of unpredictability.

Control of Conditions

A more satisfactory approach is to seek quality of evidence rather than mere quantity. Traditionally, quality of evidence has been sought by controlled

conditions, in order to reduce the effects of extraneous causes. Controlled conditions, however, are elusive. No one can ever be sure just how well-controlled they are for any one experiment. In many cases control of conditions begs the question, for control is often the very objective of the experiment. In other cases control of conditions is properly challenged on the ground that an artificial "laboratory" result will not necessarily hold true in the everyday "factory".

In field tests, such as the transportation test cited, conditions are practically uncontrolled. This is as it should be. Careful execution of the test to avoid severe handling, for instance, would itself vitiate the test, for the objective is to study conditions as nearly as possible as they will be expected in practice. In factory tests, conditions are usually somewhat controlled, for this is, or should be, the normal factory situation. In laboratory tests, conditions are usually as closely controlled as the existing facilities and abilities will permit.

These various degrees of control affect the well-known "experimental error". Other things being equal, closer control means less experimental error. This has long been recognized. It has also been recognized that experimental error tends to obscure the result. What has often been overlooked is that obscuring of a result is only a small part of the havoc which can be wrought by extraneous causes. The results of the transportation test were not obscure; they told a clear lie. The bump test result was not obscure; it was ridiculous. The result of the machine speed test was not obscure; it told a ten-thousand dollar lie.

The New Techniques

Obscurity can be reduced by control, but often at the price of using a narrow and misleading set of artificial conditions. Even close control, however, does not remove the possibility that clear lies will be told by the result. These lies can be adequately prevented only by one or both of two additional techniques. One of these is "balancing"; the other is "randomizing". By means of <u>balancing</u>, the possible effects of <u>selected</u> extraneous causes are <u>separated</u> from the effects of the cause or causes being investigated. By means of <u>randomizing</u>, the possible effects of <u>all</u> <u>other</u> extraneous causes are <u>neutralized</u>. Furthermore, these techniques accomplish these ends whatever the degree of control.

Transportation Test

Consider the transportation test. This test had been executed by shipping all of the 2000 units in one truck. All of the 1000 units shipped in standard packing, however, had been manufactured on one day; all of the 1000 units shipped in the proposed packing had been manufactured on a different day. The conclusion that the proposed packing was inferior was therefore without sound basis. The effects of any possible differences between manufacturing dates had been completely confounded with any difference in the effects of the two types of packing.

Questioning revealed also that no thought had been given to randomizing the order of loading the packages into the truck, handling them at destination, etc. Thus it was not unlikely that severity of treatment, as well as a difference between manufacturing dates, could also have been badly confounded with any effects of the different types of packing.

It is clear that, if all 2000 units had been manufactured on one date instead of two, there would have been no confounding of effects of type of packing with effects of differences between manufacturing dates. There would, however, have been a narrow base for the experiment. If a difference between manufacturing dates had existed, then conclusions valid for one date would not necessarily have been valid for another date. In any case, an experiment based on one date is incapable of shedding any light on differences between dates.

Confounding of between-date effects could have been avoided and the two dates still included in this instance by an extremely simple plan of balancing.

In each type of packing, there were 100 units to the package. There were, therefore, 10 packages of standard packing and 10 packages of proposed. A simple balanced set for the experiment would therefore consist of:

5 packages (500 units) standard packing, manufacturing date I
5 packages (500 units) standard packing, manufacturing date II
5 packages (500 units) proposed packing, manufacturing date I
5 packages (500 units) proposed packing, manufacturing date II

Such a set would have balanced the effects of manufacturing dates and types of packing, and these two effects could have been separately analyzed, each without any confounding due to the inclusion of the other.

There might, however, have been other effects within the transportation trip and within each of the manufacturing dates which could also have confounded the results.

Suppose, for instance, that all 10 packages of date I had been placed at the front of the truck and all 10 packages of date II at the rear; or, probably worse, suppose that all 10 proposed packages had been in a lower layer of the

loaded truck and all 10 standard packages in an upper layer. In either case any differences in transportation severity due to these locations would have been confounded with the effects which were being investigated.

It is clear that, if the packages had been loaded on the truck in a truly random order, the placing of all of the proposed packages in a lower layer and all of the standard packages in an upper layer would have been extremely unlikely. So would <u>any other</u> loading arrangement which could cause a severe confounding of the results, such as the placing of all proposed at the front and all standard at the rear; always stacking proposed on proposed, standard on standard; placing a large number of one type along a diagonal of the load, etc. It is impossible to name all the conceivable reasons for bad confounding, but they are <u>all</u> removed from likelihood by the simple process of true randomizing.

Again, suppose that manufacturing conditions had changed appreciably during a day; e.g., that product made in the latter part of the day was less shock-resistant than that made earlier in the day. Suppose, furthermore, that the 1000 units selected for standard packing had been made in the early part of each of the two manufacturing dates, while the 1000 units selected for proposed packing had been made in the latter part of each of the two manufacturing dates. In such a case also the experiment would have been confounded, and in the analysis this difference, really due to manufacturing conditions, would have appeared as evidence that the proposed packing was inferior.

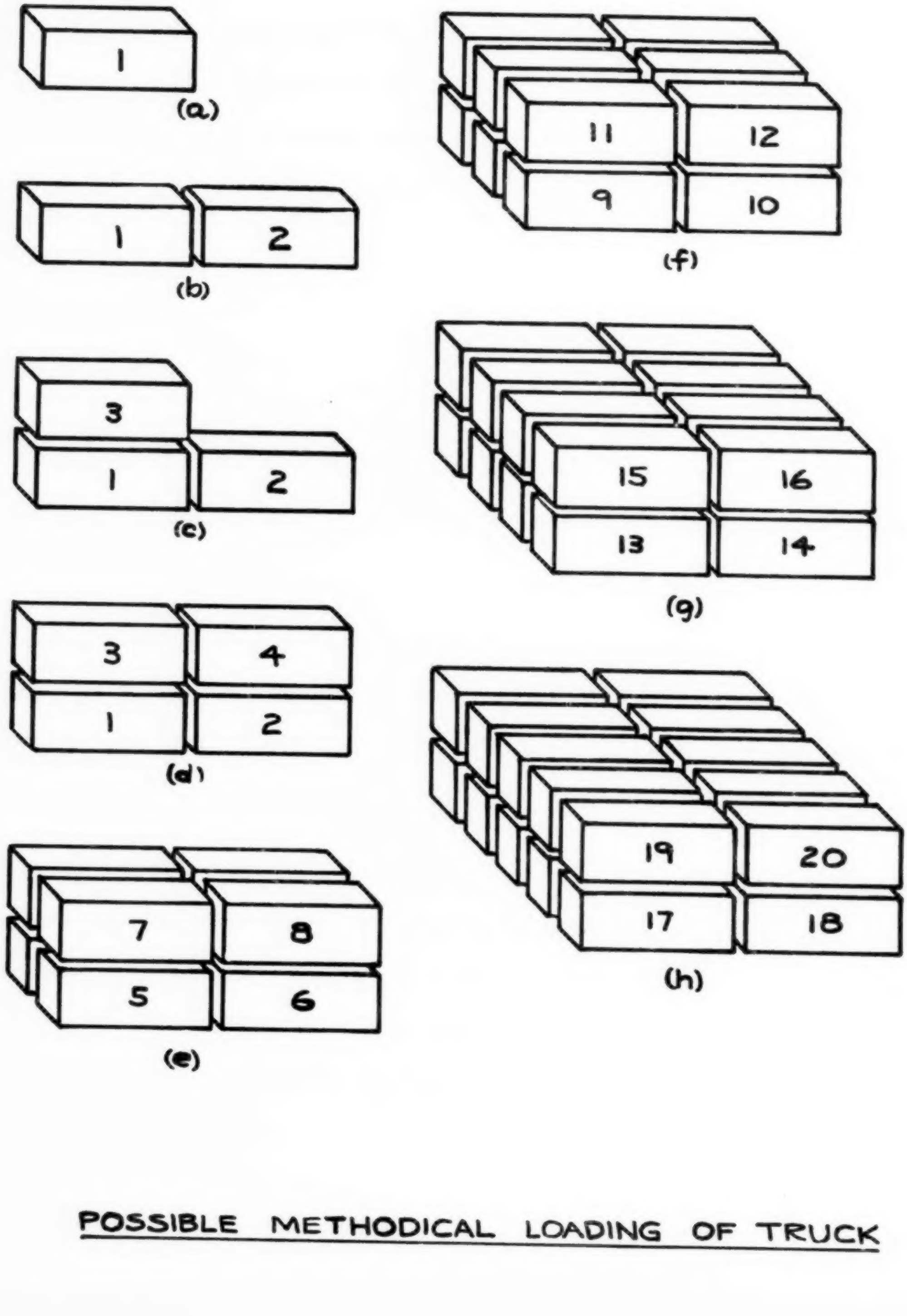
Again it is clear that, if the units for the proposed and standard packing had been selected from the production flow by a truly random method, it would have been extremely unlikely that the assignment of units to type of packing would systematically follow fluctuations in manufacturing conditions to an extent which would cause severe confounding.

The Fallacy of Alternating

Some experimenters attempt to circumvent the problem of randomizing by alternating, on the assumption that such a procedure accomplishes the same thing. Reflection, however, indicates that this is not so.

Consider, first, the truck loading. Suppose that the load of 20 packages were placed in the truck in two layers, each consisting of two rows of 5 packages. A normal way to pack such a load might be to place packages in some order such as indicated in the attached sketch. If, in this sketch, the observer is looking toward the front of the truck, the packages would be placed in this order:

First package:	front, left, lower
Second package:	front, right, lower
Third package:	front, left, upper
Fourth package:	front, right, upper
Fifth package:	second from front, left, lower
.	.
.	.
.	.
Eighth package:	second from front, right, upper
Ninth package:	third from front, left, lower
.	.
.	.
.	.
Twentieth package:	fifth from front, right, upper



POSSIBLE METHODICAL LOADING OF TRUCK

If, now, the loading is done by "alternating" in the following order:

standard date I

standard date II

proposed date I

proposed date II

standard date I

.

.

.

etc.

all of the standard packing will find its way to the lower layer, and all of the proposed to the upper; all of date I will be at the left, and all of date II at the right.

Consider, second, the selection of units for the proposed and standard packing. Suppose that units are selected as they come along a conveyor fed by a machine. Suppose further that alternating is done by assigning the first unit to proposed, second to standard, third to proposed, etc.

Such a machine might well have an even number of heads. Suppose that there are 10 heads, and that head number 7 is making a weaker product than the rest. The units made on head number 7 will all find their way into one type of packing. Which one, of course, will depend on when the selection starts. In either case, however, there would be severe confounding of this manufacturing fault with the effect of type of packing.

Planned Randomizing

Usually randomizing cannot be accomplished by the simple expedient of issuing instructions to "load in random order" or the like. If, for instance, in the truck loading case discussed, the standard and proposed packages were different in appearance or had been externally identified for the purpose of the test, the shippers would have been quite likely to load first all of one group, then all of another, etc.; for the simple reason that they would want to keep track of what they were doing.

It is far safer to do intentional and planned randomizing. In this case, for instance, a card drawing scheme would be applicable. This could be carried out by assigning to each of the four groups one of the four suits of a standard deck of cards, for instance:

standard, date I - clubs

standard, date II - diamonds

proposed, date I - hearts

proposed, date II - spades

If, now, 5 cards of each of the four suits are selected, these 20 cards could represent the 20 packages. After thorough shuffling, these cards could then be drawn in random order, and numbers assigned to the packages in the order in which the cards are drawn. During the card drawing the assigned numbers (from 1 to 20) would be clearly marked on the outside of the packages and the order of drawing recorded and preserved for subsequent identification. Then the shippers and all others who handle the product would be instructed to handle "in numerical order".

The Validity of Conclusions

After the test had been completed, the results could then be conveniently tabulated by recording the number of defective units in each of the 20 packages, tabulated by groups. This is indicated in the results of an actual test, very similar in nature, and carried out by a similar plan of balancing and randomizing. The final data were tabulated thus:

	Date I	Date II	Total
	0	0	
	2	0	
Standard	1	0	8
	1	1	
	3	0	
	2	1	
	0	0	
Proposed	2	0	10
	1	1	
	2	1	
Total	14	4	18

In this case it was nearly evident from the raw data that a real difference between manufacturing dates had been shown, while no real difference had been established between types of packing.

A statistician, called in at the end of the experiment to analyze these data, would have reached substantially the same conclusions, except that he would have been likely to express them more cautiously. He might have said, for instance, that the indicated difference (8 and 10) between types of packing could easily have occurred by chance. He would also have said that the <u>indicated</u> difference (14 and 4) between manufacturing dates is greater than could reasonably be due to chance acting alone. It would also have been apparent to him that this indicated difference could not have been due to type of packing. For type of packing was a selected cause which had been

separated by balancing. Even if a difference in effect between types of packing had existed, it could not have vitiated the comparison between manufacturing dates, because there were the same number of units of each manufacturing date in each type of packing.

The statistician's cautious statement that the "indicated" difference between manufacturing dates appears to be real; i.e., must in part at least be due to causes which did not follow the laws of chance, would have been about as far as he could go. He could not have stated that this difference was due to manufacturing conditions; i.e., existed in the product before the transportation test. For the cause "other than chance" which was responsible for the observed difference might, as far as he knows, have been due to peculiarities in the test.

But the engineer who conducted the test knew that this could not be so. He knew that his planned randomizing had made every set of causes, except two, follow the laws of chance. Those two are the types of packing and the manufacturing dates. Since the effect of type of packing had been separated by balancing, there remained only one set of causes which could reasonably have had an effect "other than chance". That cause was the difference between manufacturing dates; i.e., it had existed in the product before the test was made, and was due solely to the conditions of manufacturing.

Note that balancing works both ways. The experiment was planned to study the comparative effects of the two types of packing; the between-date difference was simply a possible extraneous effect which had been balanced to avoid confounding the main effect. As it turned out, the "extraneous" effect was the real one. This real effect was, in turn, kept free of being confounded by any difference which might have existed between the effects of the types of

packing. This would not have been so if the number of units of each date's manufacture in each of the two types of packing had been unequal.

In the original experiment, in which the two effects were confounded, and from which erroneous conclusions had thus been drawn, easy assumptions had been made. By implication at least, it had been assumed that any differences in strength of product due to time-to-time changes in manufacturing conditions, and any effects due to differences in severity of treatment, were small enough to be disregarded. But those assumptions were opinions. After the experiment had been completed, the conclusions drawn had depended on those opinions; the facts had not spoken for themselves.

Filament Test

Consider the test of filament strength and its strange result of 10 shattered bulbs with one type of tungsten wire and none with the other. The tester suggested that a chip of glass must have become embedded in the bump pad during the course of the test. The fact that this explanation was offered was itself evidence that all of the lamps of one group had been tested first, then all of the lamps of the other group. Perhaps, therefore, it was sheer good fortune that the bulbs did shatter. For if a small chip of embedded glass can have such a marked effect on bulb breakage, certainly there are countless other possible time-to-time variables in test equipment and procedures which might have affected the test of the filaments even if the bulbs had not broken. Such variables would have vitiated the results anyway, and comparisons between types of tungsten wire, the very purpose of the test, would not have been valid.

Since these time-to-time variables are entirely unknown and cannot be enumerated, their effects cannot be balanced. They can, however, be neutralized by

randomizing. In this test randomizing would have been extremely simple. If the tester had tested one lamp at a time, and determined by, say, the drawing of cards, from which group the next lamp should be selected, the test itself would have been completely randomized.

In this case, however, randomizing of the test would not of itself have provided a sound basis for valid conclusions. The test was a test of wire, and the wire had been through many processes before it had become filaments mounted inside incandescent lamps. To enumerate only a few, there were coiling, annealing, treating, sintering, mounting, sealing, exhausting, and flashing. Since any one of these processes could have affected filament strength, its effects could become severely confounded with differences in the original wire itself unless they also were neutralized.

Consider, for instance, the mounting process. Several sources of variability can be named, countless others must exist. Among the namables are the tightness of clamping of lead wires to filaments, the spacing of the supports, the tension of the filament, etc. Suppose that during the mounting of the test filaments the tightness of the clamping had been progressively increasing. If, then, all 20 filaments of one group were mounted first, and all 20 filaments of the other group afterwards, any effect of clamping tightness would vitiate any conclusions about type of wire. If, however, the order of feeding filaments to the mount machine had been randomized, any effect due to clamping variables would have been neutralized and would not have vitiated the conclusions.

Before the experiment had been started, or during its execution, it would have been easy to assume that the effects of clamping tightness were either nonexistent or small enough to be disregarded. But such an assumption would have

been an opinion. After the experiment had been completed, the conclusions drawn would then have depended on that opinion; the facts would not have spoken for themselves.

Once the specimens of wire, the difference between which was to be tested, had been selected, they should then have been sent through every subsequent process, including the test itself, in random order. It is clear, however, that the mere feeding of units to a machine in random order will randomize every step of the process which is performed by that machine; and the introduction of the units to any series of processes in random order will randomize every one of the processes, provided the units are not sorted into order between processes.

Machine Speed Test

Consider next the machine speed test. It was apparent to one and all that "something else", other than machine speed, must have entered into the experiment to give the definite evidence of longer life at the slower speed. Many are of the opinion that those "something elses" are simply the perversities of nature to which experimenters must become resigned unless they are willing to repeat their experiments several times and make no decisions unless consistent results can be shown.

To those who have formed the habit of thinking in terms of variability, however, such a result was not surprising. Every part of every process is exposed to many sets of causes of variability at all times, and it would be fantastic to expect all of the rest to behave themselves while a selected one is varied. The experiment was therefore repeated, this time using the randomizing techniques.

The first step was to randomize the mounts to be fed to the sealing machine whose speed was to be varied. The mounts were the product of the last process performed before sealing. The randomizing of mounts, therefore, randomized the effects of every process which preceded sealing, and effects due to sealing would thus be unlikely to be severely confounded with effects due to any process which preceded it.

The random order assigned to the mounts was then maintained through every process subsequent to sealing, including the test itself. Thus all sets of causes, both before and after sealing, affected the lamps in random fashion, and the only causes which could show a real effect in the results would be those due to machine speed itself.

Analysis of the final results showed that machine speed was without measurable effect on life. Further ferreting finally disclosed the real cause in an entirely different part of the process. There is every reason to believe that, if adequate precautions had been taken to neutralize extraneous causes in the original experiment, erroneous evidence of an effect due to machine speed would have been lacking. For in that first experiment the easy assumption had been made that the effects of all other extraneous causes were either non-existent or small enough to be disregarded. But such an assumption was an opinion. After the experiment had been completed, the conclusions drawn had depended on that opinion; the facts had not spoken for themselves.

Special Designs of Experiments

Much has been written and said about factorial experiments and their value in the study of several sets of causes simultaneously. By means of such balanced designs of experiments it is possible to increase both the breadth of the base of an experiment and its efficiency in preventing duplication of experimental

effort. This experimental efficiency is even more spectacular in the designs known as Latin Squares and Graeco-Latin Squares.

Experience indicates, however, that too little emphasis has been placed on the importance of randomizing the effects of extraneous causes which are always present, even in these ingeniously balanced arrangements.

Consider a comprehensive experiment being conducted at the present time to study the effects of various cathode design parameters on the life of fluorescent lamps. The cathode is a tungsten filament of the "coiled coil" type. It is manufactured by winding tungsten wire on molybdenum mandrel wire, then winding this combination again on a steel mandrel pin which is automatically withdrawn when the filament is cut to length. The molybdenum mandrel is then removed by acid treatment. The filaments are manufactured in "schedules" of several hundred or more at a time.

The five parameters whose effects are being studied are:

Tungsten wire diameter

Molybdenum mandrel diameter (Primary mandrel)

Turns per inch wound on molybdenum mandrel (Primary T.P.I.)

Steel mandrel pin diameter (Secondary mandrel)

Turns per inch wound on steel mandrel pin (Secondary T.P.I.)

By selecting five different values of each of the five different parameters, a 5 x 5 Graeco-Latin Square was laid out to study these five parameters as thoroughly as could be done with a test on only 25 lamps.

After the plan of the experiment had been arranged, an order was transmitted to the Filament Department and attached was a list of the 25 different designs of filaments for the experiment. The list was made out as shown below,

where the five values of tungsten wire diameter are represented by A, B, C, D, E; the five values of primary mandrel diameter by I, II, III, IV, V; the five values of primary T.P.I. by 1, 2, 3, 4, 5; the five values of secondary mandrel diameter by a, b, c, d, e; and the five values of secondary T.P.I. by v, w, x, y, z.

Design No.	Wire Diam.	Primary Mandrel Diam.	Primary T.P.I.	Secondary Mandrel Diam.	Secondary T.P.I.
1	A	I	1	a	v
2	A	II	2	e	x
3	A	III	3	d	z
4	A	IV	4	c	w
5	A	V	5	b	y
6	B	I	2	b	w
7	B	II	3	a	y
8	B	III	4	e	v
9	B	IV	5	d	x
10	B	V	1	c	z
11	C	I	3	c	x
12	C	II	4	b	z
13	C	III	5	a	w
14	C	IV	1	e	y
15	C	V	2	d	v
16	D	I	4	d	y
17	D	II	5	c	v
18	D	III	1	b	x
19	D	IV	2	a	z
20	D	V	3	e	w
21	E	I	5	e	z
22	E	II	1	d	w
23	E	III	2	c	y
24	E	IV	3	b	v
25	E	V	4	a	x

The combinations of designs in this list met the requirements for a Graeco-Latin Square, in that each value of each parameter was associated with each value of each of the other parameters once and only once.

Made out in this order, however, the list would have implied that the schedules of experimental filaments should be manufactured in the listed order. If this

had been done, there would have been serious faults in the procedure. With respect to wire diameter, there would have been the fallacy of manufacturing all schedules of filaments with wire diameter A, then all with wire diameter B, etc. In this way any possible time-to-time trends due to extraneous causes would have been severely confounded with the effects of wire diameter. With respect to primary mandrel diameter, and to some extent with respect to the remaining three parameters, there would have been the fallacy of alternating. In this way, any possible cyclical variations, keeping in step or nearly so with the alternations, would have caused severe confounding.

The order of the list was therefore permuted, by card drawing, and the revised list as sent to the Filament Department was as shown below.

Design No.	Wire Diam.	Primary Mandrel Diam.	Primary T.P.I.	Secondary Mandrel Diam.	Secondary T.P.I.
1	C	II	4	b	z
2	C	III	5	a	w
3	B	III	4	e	v
4	B	V	1	c	z
5	D	IV	2	a	z
6	A	IV	4	c	w
7	E	I	5	e	z
8	E	V	4	a	x
9	A	III	3	d	z
10	B	IV	5	d	x
11	A	I	1	a	v
12	E	II	1	d	w
13	A	II	2	e	x
14	D	III	1	b	x
15	B	II	3	a	y
16	A	V	5	b	y
17	D	II	5	c	v
18	E	IV	3	b	v
19	E	III	2	c	y
20	D	V	3	e	w
21	C	V	2	d	v
22	B	I	2	b	w
23	C	IV	1	e	y
24	C	I	3	c	x
25	D	I	4	d	y

This included the same designs as the list which had been laid out in order, but the suggested procedure had been completely randomized. In addition to revising the list, specific instructions were issued directing that the schedules of filaments be manufactured in the order shown, and the job was followed closely to make sure that the listed order was followed to the letter at every stage of the filament manufacturing process.

It was fortunate that this was done, for the following of random order often adds to the work involved in carrying out certain parts of a process, and there is a temptation to sort, in order to expedite the job. This case was no exception.

For instance, between primary and secondary coiling, the tungsten-molybdenum combination is annealed in a hydrogen furnace. The specified temperature of the furnace is different for different values of primary mandrel diameter. Furthermore, it takes time to change the furnace temperature. In order to expedite the experimental procedure at this point, the foreman proposed that he use five different furnaces, setting one at the temperature for diameter I, a second at the temperature for diameter II, etc., and then run all diameter I through the first furnace, all diameter II through the second, etc. The engineer, of course, could not allow this procedure. The foreman than asked if, using a single furnace throughout, he could first set at the temperature for diameter I and anneal all five schedules of diameter I, then set at the temperature for diameter II and anneal all five schedules of diameter II, etc. Again the engineer refused, and insisted that the listed order be followed.

This insistence was absolutely necessary. There can be differences, other than temperature, between furnaces and within one furnace from time-to-time. Among the possibilities are atmospheres, temperature gradients, etc. The

effects of such variables belong in the experimental error, and must not become associated with a parameter being investigated. Randomizing keeps them in the experimental error. The procedures proposed by the foreman would have made them become associated with primary mandrel diameter.

The objection might be made here that, since different primary mandrel diameters require different annealing temperatures, this temperature effect is confounded with primary mandrel diameter anyway. Reflection indicates, however, that specified changes in annealing temperature are not extraneous causes, but are <u>necessarily</u> associated with changes in mandrel diameter, and decisions between mandrel diameters necessarily carry with them decisions between corresponding annealing temperatures. The purpose of the random procedure is to prevent confounding by neutralizing the effects of <u>unknown</u> and <u>uncontrolled</u> causes which otherwise might become associated with the parameters during the experiment, but would not necessarily be so associated at other times.

The lamps, now on test, will not reach the ends of their lives until well over a year from now. During the experiment, it would have been easy to assume that these extraneous between-furnace and time-to-time effects were either non-existent or small enough to be disregarded. But such an assumption would have been an opinion. When the data have been assembled, the conclusions then drawn would depend on that opinion; the facts would not have spoken for themselves. Erroneous conclusions, after many months of experiment and test, would be a high price to pay for a few hours saved during experimental filament processing by work simplification which "derandomizes" the procedure.

Balancing Vs. Randomizing

Frequently there is a clear choice between balancing and randomizing a particular source of variability. Consider, for instance, a recent experiment on incandescent lamps.

The objective was to study the effect of varying degrees of filament distortion on the life of lamps, and also how this varied between two different types of mount construction. Two lamps of construction I and two lamps of construction II were selected for each of five different degrees of filament distortion, making 20 lamps in all.

At the life test laboratory there were two test racks, each having 10 sockets. Any difference which might exist between racks was therefore a source of variability. Such a source of variability would be badly confounded with type of lamp construction if all 10 lamps of construction I were tested on rack A and all 10 lamps of construction II were tested on rack B. It would be better to neutralize this possible effect by assigning lamps to racks at random. Would it be better still to balance this possible effect by assigning one lamp of each construction-distortion to each rack?

In the actual test, the latter course was chosen; i.e., the possible effect of the racks was balanced. The final life data (expressed as a percentage of rated life) were as shown below.

		Construction I		Construction II		
		Rack A	Rack B	Rack A	Rack B	Average
	1	116	115	129	113	118
Degree of	2	94	83	106	134	104
Filament	3	81	93	135	98	102
Distortion	4	118	109	105	129	115
	5	80	73	102	87	85

Averages:	Rack A:	107	Construction I:	96	105
	Rack B:	103	Construction II:	114	

Analysis of these data indicated that the difference between constructions is highly significant, the differences due to distortion are significant, but that there was no measurable difference between the racks.

If, however, there had been a real difference between the racks, the data might well have appeared thus:

		Construction I		Construction II		
		Rack A	Rack B	Rack A	Rack B	Average
	1	124	107	137	105	118
Degree of	2	102	75	114	126	104
Filament	3	89	85	143	90	102
Distortion	4	126	101	113	121	115
	5	88	65	110	79	85

Averages:	Rack A:	115	Construction I:	96	105
	Rack B:	95	Construction II:	114	

Analysis in this case would indicate a real difference between racks, and would also indicate, as before, a high significance in the difference between constructions, and significant differences due to distortion. Thus, even with a real difference between racks, the value of the balanced experiment would be unimpaired.

Suppose, on the other hand, that with this real difference between racks existing, the test had been randomized instead of balanced. If this had been done,

each lamp could have been placed on either rack, and the life data might well have appeared thus:

		Construction I	Construction II	Average
Degree of Filament Distortion	1	127 104	105 117	113
	2	102 95	126 114	109
	3	89 105	123 90	102
	4	101 126	121 93	110
	5	88 65	110 99	90
	Averages	100	110	105

Analysis in this case would indicate that the difference between constructions <u>and</u> the differences due to distortion, both shown to be significant in the balanced experiment, would appear to be within reasonable limits of chance and therefore not considered significant at all.

The reason for this is the loss of sensitivity due to randomizing the racks. By randomizing racks, an appreciable source of variability would have been thrown into the experimental error, and this larger experimental error would have made the real effects obscure. In the experiment in which the effect of the racks was balanced, the difference between racks was stripped out of the experimental error, and this made the experiment more sensitive.

It is apparent that nothing is lost in any case by balancing, whether the potential cause has a real effect or not, but that sensitivity can be lost if a real cause is randomized. When a clear choice is available, it is better to

balance a potential cause which could have been randomized than to randomize a real cause which should have been balanced.

The Problems of Application

The applications of these techniques of balancing and randomizing present to the Quality Control Engineer a most interesting and challenging opportunity. While the principles are universal, each actual case has its own peculiar physical problems which tax the ingenuity. The greatest of the problems, however, are the human ones. The perversity of inanimate objects can be neutralized, but the fallibility of human beings will present difficulties for many years. One of these is the tendency persons have toward order. This is natural, for most tasks are best accomplished when carried out in an orderly manner. In investigations of the nature we have been discussing, however, order can be carried too far. This will be so when operators and testers "derandomize" the procedures by sorting. It has been found necessary to take special precautions to prevent this. Careful instructions and close follow-up are essential.

A more difficult problem in the human relations involved is the convincing of various persons that careful randomizing is actually a necessary prerequisite to valid conclusions. Generally the more familiar a person is with a process, the more difficult it is to convince him of the necessity of randomizing at certain stages of his particular process. His tendency is to assume that there are stages at which randomizing is unnecessary. If a specific potential extraneous cause is suggested to illustrate the need of randomizing, he either suggests a specific remedy in the procedure for <u>that</u> cause, or dismisses the cause as unimportant. As other potential causes are suggested, he continues to suggest specific remedies or to dismiss the causes as unimportant.

The suggesting of specific remedies for specific causes amounts to the classical "control of conditions". The specific remedies often add substantially to the technical problems and cost of the experiment, and they can never give full confidence in the validity of conclusions, for it is never possible to name <u>all</u> the possible extraneous causes. Randomizing, on the other hand, even if it seems to be "fussy", is usually basically simple, and does yield full confidence, for it automatically neutralizes <u>all</u> extraneous causes.

Dismissing a suggested cause, or other causes, as unimportant amounts to injecting an opinion into the experiment before it is performed. The fundamental purpose of an experiment is not to confirm opinions, but to make the facts speak for themselves in the language of data. If it is worth while to invest time and money in experiments, it is certainly worth while to make the data tell the truth.

NO. 8
Price 25¢

ACCEPTANCE INSPECTION

by

A. C. Richmond
General Supervisor
Inspection Methods Research
International Harvester Co.
Chicago, Illinois

for presentation at the

FOURTH NATIONAL CONVENTION
and
FIFTH MIDWEST CONFERENCE

of the

AMERICAN SOCIETY for
Quality Control

June 1 and 2, 1950
Milwaukee Auditorium, Milwaukee, Wisconsin

FIFTH MIDWEST QUALITY CONTROL CONFERENCE
P. O. Box 1204, Milwaukee 1, Wisconsin

ACCEPTANCE SAMPLING

A. C. Richmond
General Supervisor, Inspection Methods Research
International Harvester Company, Chicago, Illinois
Fellow, Illinois Section, A.S.Q.C.

I wonder if many of us realize the part that sampling inspection plays in the mass production of material. Do we realize that technically "100% inspection" simply does not exist?

Let us consider for example the production and fabrication of steel. The chemical properties of a 100-ton or larger heat of open hearth steel are determined on the average by the analysis of less than 2 ozs.

After a heat of steel is processed into bars, shapes, etc., at the mill, it is shipped to a customer for forging, heat treating, machining, etc. If the analysis of the steel is of importance to its processing or use, it is usually checked for chemical analysis at the customer's plant where again a very small sample is used. However, the total amount of material actually inspected for chemical analysis ordinarily is less than one part in two million.

Many of the important steel parts that are used in modern equipment are heat treated; and we have specifications for their tensile strength, elongation, hardness, etc. Consider an important part like a connecting rod bolt in an internal combustion engine. Many plants will tell you that they check the heat treatment of these bolts 100%. But do they? In most cases, you will discover that they check the hardness of these bolts either by a Brinell hardness check on the head or a Rockwell hardness check on the side. In the case of the Rockwell check, compare the area actually measured with the total area of the bolt and again we find that the sample inspected is very small compared to the universe from which it is selected.

When we inspect parts for size, a similar condition exists. Let us consider the classification of piston pins for size. It is customary to check two or three diameters on each pin; and, from this small sample of the infinite number of diameters which are present in every pin, we make a decision as to which classification it belongs.

I mention the above examples in order to bring out the fact that we in industry, whether we realize it or not, are constantly making decisions based on samples and, in most cases, these samples are small in relation to the quantity of material under investigation. Since most of our decisions are based on sampling, it seems obvious that we should all be interested in methods that tend to give us the maximum amount of information from these small samples.

If I understand statistics correctly, this maximum amount of information from a given sample is one of the important jobs on which our statisticians are constantly working. Among their major contributions to industry are the various series of acceptance sampling tables which are becoming more and more generally used by our government and industry.

My first real experience with such tables was during the last war. As chief of the inspection branch of the Chicago Ordnance District, on leave from International Harvester, it became my responsibility to supervise the introduction and use of army ordnance tables in the acceptance of ammunition materiel. A short time later, these same tables were used to a varying degree in other branches of the procurement services. Prior to the introduction of Quality Control in the ammunition branch, a lot of data were collected and what was believed to be reasonable acceptable quality levels were established for many of the ammunition components.

Our largest suppliers were visited. Standard sampling tables and pro-

cedures were explained to them. On a set date, we started to use Quality Control Acceptance Sampling on a major part of the materiel we were accepting for the ammunition branch in the Chicago Ordnance District. That was mistake number one; we "bit off more than we could chew." We learned the hard way some of the reasons why Quality Control should be introduced slowly. Within a few days after we started using army ordnance tables, the rejections were so high that shipments were very seriously curtailed and, in some cases, completely stopped. Since we had been in the war for many months and this materiel had been found to be reasonably satisfactory, it was obvious that something was wrong with this new inspection system. Orders were to return to the old method until the trouble was found and corrected. It was soon found that the acceptable quality levels were set too high. I do not believe that this condition was directly the fault of those who set these levels, but rather to a chain of events.

Prior to the adoption of quality levels, a survey was made in the plants of our better producers to determine their process averages. The resident ordnance inspectors were asked to make a record of the number of pieces inspected and number rejected for a given length of time and report the process average.

Now here is what happened. Many of these operations were found to have lower process averages than had been thought. When results of the various surveys were received at Washington, I have been informed that in the interest of better quality they only considered the process average of better producers and so came up with a set of standards only a few could meet.

New surveys were quickly made, more realistic acceptable levels were set

up in accordance with instructions that later became a part of Ordnance Inspection Manual M608-8.

I quote from page 7, paragraph 18, of the February, 1944, issue. "The choice of an acceptable quality level must be based on consideration of two factors: first, the quality that is necessary for the proper functioning of the materiel; and second, the quality that can be reasonably expected from the majority of reliable producers. A balance must be struck between these two considerations, and this balance must realize the necessary quantity of usable materiel from the producing sources which are, or can be made, available."

As the war advanced and materiel became more and more scarce, we realized the importance of "...this balance must realize the necessary quantity of usable materiel..." Since the war, I am afraid that many times some of us have been prone to think in terms of the last phrase only.

As soon as these army ordnance tables were being used successfully, the quality of material arriving at loading plants, depots, and overseas improved so much that we received many favorable comments. We were able to cut our inspection force one-third during the time that the amount of materiel accepted increased nearly fifty per cent.

I have mentioned this experience in order to emphasize the importance of the fact that the adoption of a realistic acceptable quality level is necessary for the successful use of any of our acceptance tables. Unfortunately you will not find this level worked out for you in any of the literature. I suggest that you first determine the process average of the material you are now using and, unless it has proven to have been very unsatisfactory, use it at least temporarily.

There are several other points that must be given consideration in order

to realize the most from the use of Quality Control sampling tables. One is to distinguish between inspection and sorting. To me, inspection is the act of determining whether or not material is satisfactory to use as produced, or presented for use. If it is found to be unsatisfactory, it can often be made satisfactory by sorting or some other form of rework.

Somewhere in the instructions on the use of sampling tables, you will usually find words to the effect that such and such is so if the rejected lots are "detailed." Believe me, this is one part of the "fine print in the contract" you had better read. Why the word detail is used, I do not know, for I am sure they mean what we here in the midwest call screening or sorting. It is necessary to have a clear definition of inspection and sorting before many questions frequently asked concerning Quality Control can be answered.

One such question is, "How much inspection will this plan require?" Another is, "What will be the quality of outgoing material which has been submitted to this plan?" Or, in Quality Control terms, "What will the Average Outgoing Quality Limit be?"

If we distinguish between inspection and sorting, the first question is easily answered. If the rejected lots are sorted 100% and defective items replaced, answers to the second question can be read directly from our tables.

Do you realize that the average outgoing quality of anybodies' product is simply their process average, improved by removing defective material either by sorting or reworking in some other manner. I firmly believe that, when this fact is understood, much of the mystery surrounding sampling inspection is cleared up. You then grasp what is meant by the often heard expression, "Quality cannot be inspected into a product."

When you read the literature on sampling inspection, you soon become

aware that the principal effect of such inspection is to force the producer to maintain a satisfactory process average. Improvement of outgoing quality simply by removing defects in the samples is usually very small. Unless the quality of his product is very bad, a producer can soon dispose of all of his material simply by resubmitting rejected lots.

In the book, "Sampling Inspection," written by the Statistical Research Group, Columbia University, the authors point out that even if a lot of material were so poor that it would be rejected fifty per cent of the time, the odds are 8 to 7 that it would be accepted if submitted three times.

Since we have been emphasizing some of the shortcomings, you may well ask, "Why use Statistical Quality Control?" There are many excellent reasons. There is a risk of making wrong decisions on the quality of a lot of material when the decision is based on a sample. We pointed out earlier in this talk that nearly all inspection is necessarily performed on a sample basis. Hence, a risk of error is nearly always present in any inspection decision. What is this risk of accepting bad material and rejecting good material? When you use statistical methods, these risks can be closely calculated. The amount of inspection necessary to assure you that any given risk is not exceeded may be determined. You are then in a position to approach more intelligently the economical balance between cost and perfection.

Another important reason for using statistical tables is that we soon learn protection is not based on per cent inspection, but is closely related to the size of the sample. This is a hard fact for the "ten percenters" to grasp, and I was one of them until a few years ago. When we consider that the information on which a decision is based is contained in the sample and that the size of the lot from which it was taken does not change the amount of information in the sample, it is easy to recognize the fallacy of per-

centage inspection.

The question is sometimes asked as to why it is necessary to have so many tables. Before any table is selected, we should be able to answer two questions.

1. What is the worst quality that we can afford to reject a small per cent of the time?
2. What is the worst quality that we can afford to accept a small per cent of the time.

In industry, acceptable quality levels vary and the necessary risk of wrong decisions vary. For each combination of AQL and risk, a different plan is required, hence the need for many tables. Series of tables that cover nearly all ordinary requirements have been computed and published. A series of tables now rapidly gaining in use are the new "JAN" (Joint Army Navy) tables. The original Dodge-Romig tables and Army Ordnance tables continue to be very generally used.

This discussion leads us up to "Operating-Characteristic Curves." For every sampling plan, a curve can be drawn which shows the per cent of lots that will be accepted or rejected, on the average, for any average per cent defective of submitted lots. I have here (Slide No. 1) a characteristic operating curve selected from the Navy Inspection Manual Appendix X. This curve is drawn for a single sampling plan using a sample size of 75, an acceptance number 2, and a reject number 3, an AQL of 0.65 to 1.2 and an AOQL of 1.5 to 2.5. To the uninitiated, this curve may look complicated; but with a few definitions and a little explanation, it is quite easy to read and understand.

AQL, "Average Quality Level," in these Navy tables is that quality expressed in per cent defective which we are willing to accept 95 per cent of

the time. The numbers on the left of this curve denote the percentage of times a submitted inspection lot will be accepted. The numbers across the bottom represent the percentage of defective items in submitted inspection lots.

If we enter this chart from the left at 95, extend across to the intersection with the curve, and then drop down to the bottom line, we find the nearest per cent defective to be 1. This means that, if the average per cent of defective items in the inspection lots submitted is 1, approximately 95 per cent of the lots will be accepted.

Now suppose that we want to know how bad the lots would have to be before we would reject nine out of ten of them. We simply enter the table from the left at 10, proceed across to the curve and down to the bottom line where we find the per cent defective to be about 7. This inspection plan will therefore reject on the average 9 out of 10 inspection lots submitted which contain on the average 7 per cent defective items. In the same manner, for any average lot quality, the per cent of lots accepted or rejected can be determined.

Provided all rejected lots are sorted and the defective pieces are replaced with good ones, it is easy to determine the Average Outgoing Quality (AOQ) and Average Outgoing Quality Limit (AOQL) from the Operating Characteristic curve. By AOQL is meant the worst average per cent defective quality of outgoing product that could pass through, regardless of the quality of the lots submitted for inspection.

The Average Outgoing Quality for any average per cent defective submitted lots can be determined as follows: As an example, we will take one that is easy to figure. Suppose that under this plan a producer is submitting material that averages 3.25 per cent defective; from the curve we de-

termine that fifty per cent of his lots would be rejected. If all defective material is removed from one-half of the lots and replaced with good parts, one-half of our outgoing material will be 3.25 per cent defective and the other half will contain no defective items; therefore, our outgoing quality would be approximately 1.6 per cent defective. Similarly, the outgoing quality can be determined for any other average per cent defective of submitted lots.

Under any inspection plan, provided the inspector makes no errors, all lots will be accepted when the material is 100 per cent good and all lots will be rejected when the material is 100 per cent bad. In between these two extremes, a sampling plan will accept less and less material as the product becomes progressively worse.

From the Operating Characteristic curve, another curve called the Average Outgoing Quality (AOQ) curve can be plotted. This curve is obtained by using the outgoing quality levels and incoming percentages defective as co-ordinates. This curve starts at 0 per cent defective, rises rapidly to a peak, and then falls away rapidly until near 100 when it flattens out and reaches the base line at almost a tangent. The peak of the curve is the AOQL. I mention this curve as a means of taking some of the mystery out of one of the terms frequently used when discussing statistical sampling plans.

There are many interesting curves and formulae used in Statistical Quality Control; and, to those who like mathematics, it offers a challenge to both the professional statistician and the high school graduate.

During the last of the war, several International Harvester men attended some one of the eight-day courses in Quality Control which were authorized by the Office of Production Research & Development of the War Production Board. The successful results of Quality Control applications in two or three of our

plants attracted the interest of our top management.

Upon my release from the Ordnance Department late in 1945, I was given the job of making recommendations in regard to the general use of Statistical Quality Control in International Harvester Company plants. The first thing a study of the problem brought out was the need for an educational program.

The following plan was submitted to, and approved by, our top management:

1. A two-day meeting attended by all our Works Managers be held on Statistical Quality Control.
2. A ten-day intensive course in Statistical Quality Control be held for Chief Inspectors and other selected men from each of our Works.

Dr. Lloyd A. Knowler consented to assume the responsibility of preparing the course, and with the aid of Drs. Irving Burr, Edwin Olds, and Mason Wescott presenting it to the class. Professor Knowler and Professor Olds did a splendid job of selling Quality Control to the Works Managers with the result we had a class of 70 selected men on hand for the ten-day course held in August, 1947.

Also present during the ten-day course was one man from the Accounting Department of each of our divisions representing the Division Comptroller. These men have been of much value when it comes to presenting to management the savings resulting from Quality Control. With the four professors we just mentioned instructing our class, it is needless to say that it was very successful. Many of the men from this class have become Quality Control Engineers and fill other key positions in our Quality Control program.

In most of our plants, Quality Control is under a "Quality Control Engineer" who reports directly to the Chief Inspector. In most plants, we now have a "Quality Control Action Committee." The Chairman of this committee is usually a Staff Assistant reporting directly to the Works Manager who is

the highest official of the plant. The committee, in addition to the chairman, generally consists of the Chief Inspector, Quality Control Engineer, Planning Engineer, and Master Mechanic. In this group are the men who can take immediate corrective action when required. These "Action Committees" have proven to be one of the most valuable contributors to our Quality Control program.

In many of our International Harvester Company plants, we are using statistical sampling tables for inspecting material between departments and on inter-works shipments, as well as on material received from outside vendors.

It so happens that our plant here in Milwaukee was one of the first International Harvester plants to become interested in Statistical Quality Control. Nearly all material produced in this plant is now being inspected in accordance with standard inspection tables. Rejected lots are returned to the department at fault for rework or sorting. The results are very gratifying.

The next two slides are reproductions of actual charts which are typical of the "before" and "after" effects of statistical sampling. This first chart (slide No. 2) is a per cent defective (p) chart. It shows the per cent of defective items found in samples at the time statistical sampling was started on nozzle valve guides used in our Diesel engine injection pumps. These pumps are built at our Milwaukee Works.

Those of you who are familiar with Quality Control will notice that this same chart form can be used for $\bar{X}$-R (average and range) charts. In the space ordinarily used for plotting ranges, Milwaukee Works lists the characteristics checked. Opposite each characteristic is shown the number defectives found

in each sample. In the upper chart, the per cent of defectives found to the number of pieces in the sample is plotted.

The sub-group size is 700 which corresponds to one shift's production. The sample size is 75, acceptance number 2, rejection number 3. You recall that this is the same sample size and the same accept and reject numbers used in the operating characteristic curve we were just discussing; therefore, the AQL and AOQL are the same. In use, the shifts are designated by different colors. On these slides, they are shown as indicated.

Prior to the time this chart was started, these nozzle valve guides were 100 per cent screened by the Inspection Department. Rejections were high, resulting in a considerable amount of scrap and rework loss. In about three weeks' time, you will note that a big improvement was made in their process average. The process average was still not satisfactory, resulting in the rejection of too many lots. Examination of the chart shows that if the runout specification could be held, more than 50 per cent of the defectives would be eliminated. This problem was discussed with the design engineers, resulting in a change in the specifications from .002" to .003" permissible runout. This change you see was noted on the chart January 17.

A real effort was then made to bring the other characteristics under control. The results are shown on the next chart (slide No. 3). Note that a large majority of the samples are now free from any defect, and it is seldom necessary to reject a lot. Results: better quality, less scrap, less rework and inspection cost.

The next slide (slide #4) shows crank-cases for our Silver Diamond engine being inspected. The Silver Diamond engine is built at our Indianapolis Works and is used principally in the smaller line of International trucks.

These cases are checked for 22 characteristics in accordance with a Dodge-Romig double sampling table for 2 per cent AOQL. The table used is for lots of 51 to 100 and calls for a first sample of 23, a second sample of 23, with an acceptance number of 0 and a rejection number of 2. On account of the large number of characteristics checked, this is probably the most difficult item now being inspected in accordance with standard sampling procedure in our company.

They have an interesting method of making sure that the sample is chosen from each lot in a random manner. The crank-cases are numbered from 1 to 100 as they are received at the inspection bench. Four cards have been prepared, each card has 23 random numbers from 1 to 100. When a new lot starts to arrive, the cards are shuffled and two are drawn. The numbers on the first card become the first sample; and the second card, the second sample. Our conveyors are so arranged that the first samples pass the inspection benches immediately, and the second sample is shunted onto another conveyor. The remainder of the lot is held until final disposition of the lot is made.

The results have been exceptionally good. Under the old so called "100 per cent" inspection method, about 5 per cent of the crank-cases caused trouble on the assembly line. Now less than .5 per cent cause trouble. For approximately the same production, 17 inspectors were required under the old plan; now only four are required and, by simply increasing the lot size, it is estimated these same four men will be able to handle a 50 per cent increase in production. Many other items at this Works are now being inspected according to standard sampling plans. We are informed that Indianapolis Works expects to use them almost 100 per cent in the near future.

This Works has adopted a novel plan of grading its vendors. Nearly

all incoming material is inspected on an acceptance sampling basis, and the vendors' process average is computed for both major and minor characteristics on the basis of the samples of their product inspected each month. They are then given a "Quality Rating." This rating is made in accordance with a key sheet which is reproduced on the next slide (slide No. 5). Notice that the first eight ratings are in equal increments and the last includes all over the eighth.

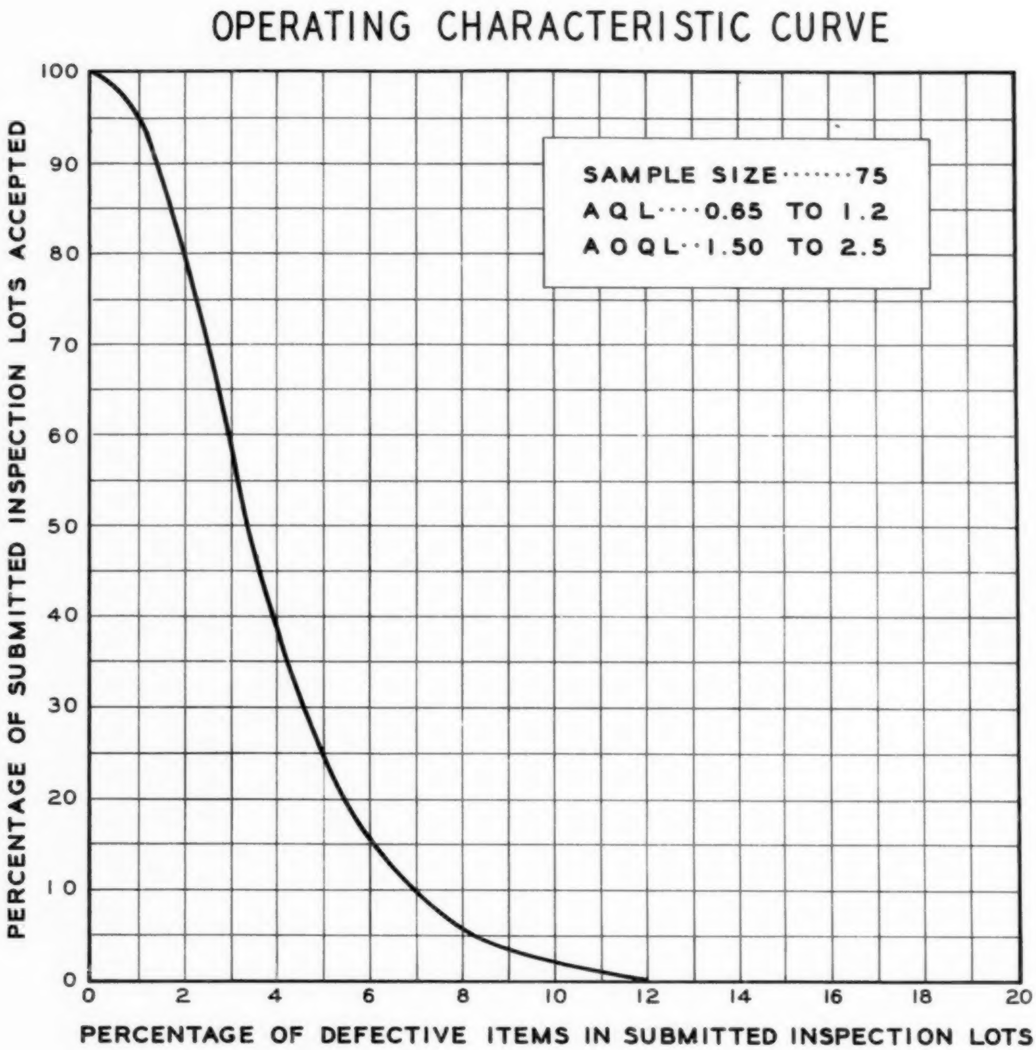
Each vendor is notified of his monthly Quality Rating. In addition, on a large board in the inspection office, each vendor's previous monthly rating is posted. You can easily guess that this board has become of interest, especially to competing vendors. The results of standard sampling and vendor quality rating have been very good. The composite ratio of defective items received decreased approximately 20 per cent during the first six months and the trend is still downward.

Statistical Quality Control has two main divisions: "Process Control" and "Acceptance Sampling." Acceptance sampling is divided into two groups: inspection by "Variables" and inspection by "Attributes." The first is a method of measuring and analyzing the variation of items in samples and between samples. In inspection by attributes, each item is classed as acceptable or not acceptable without regard to degree of non-conformance; for example, items checked with a "go, no go" gauge. This talk has been limited to a discussion of inspection by attributes.

In the International Harvester Company, we are using all of the commonly used Statistical Quality Control Techniques. In the fiscal year 1949, Quality Control was credited with being one of the major contributing factors which not only resulted in a major improvement in the quality of our product,

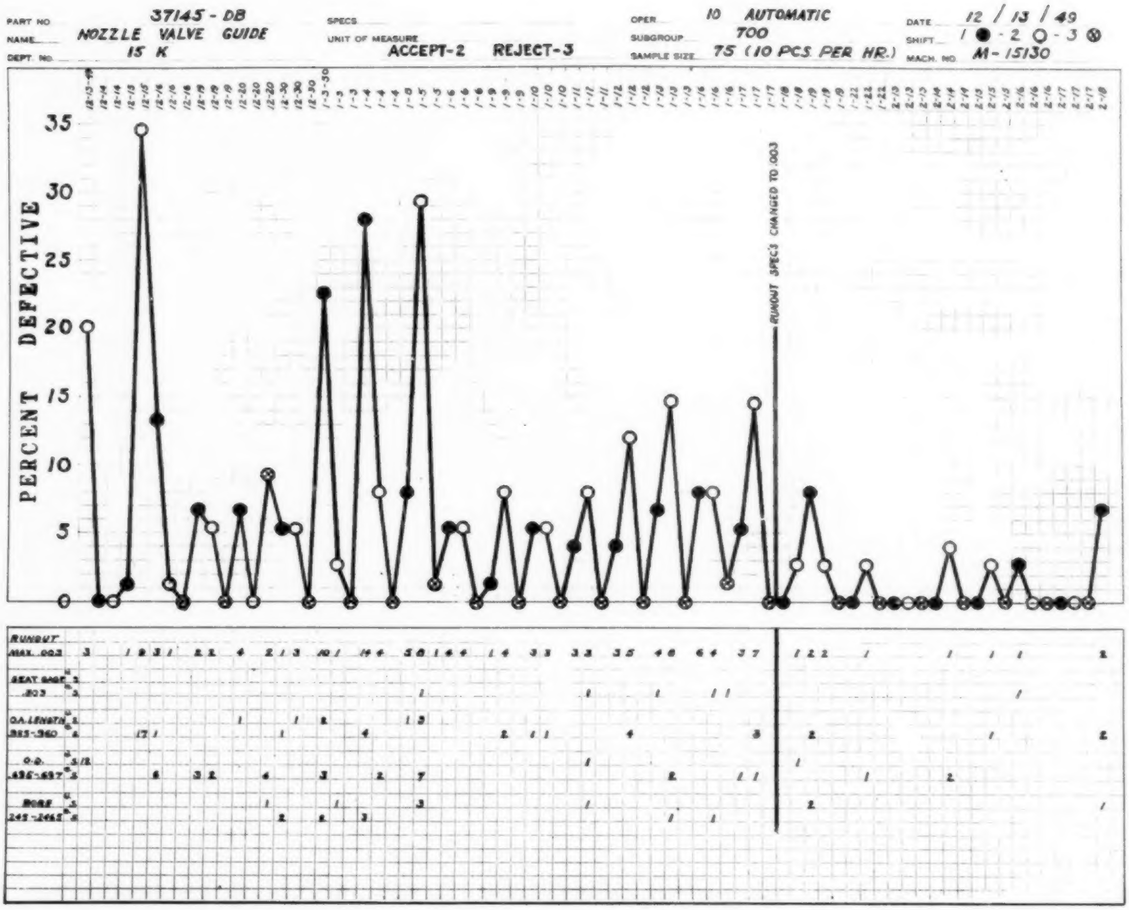
but also resulted in a savings in scrap, rework, and inspection cost of over two million dollars. According to present trends, a similar amount will be realized this year.

SLIDE NUMBER 1

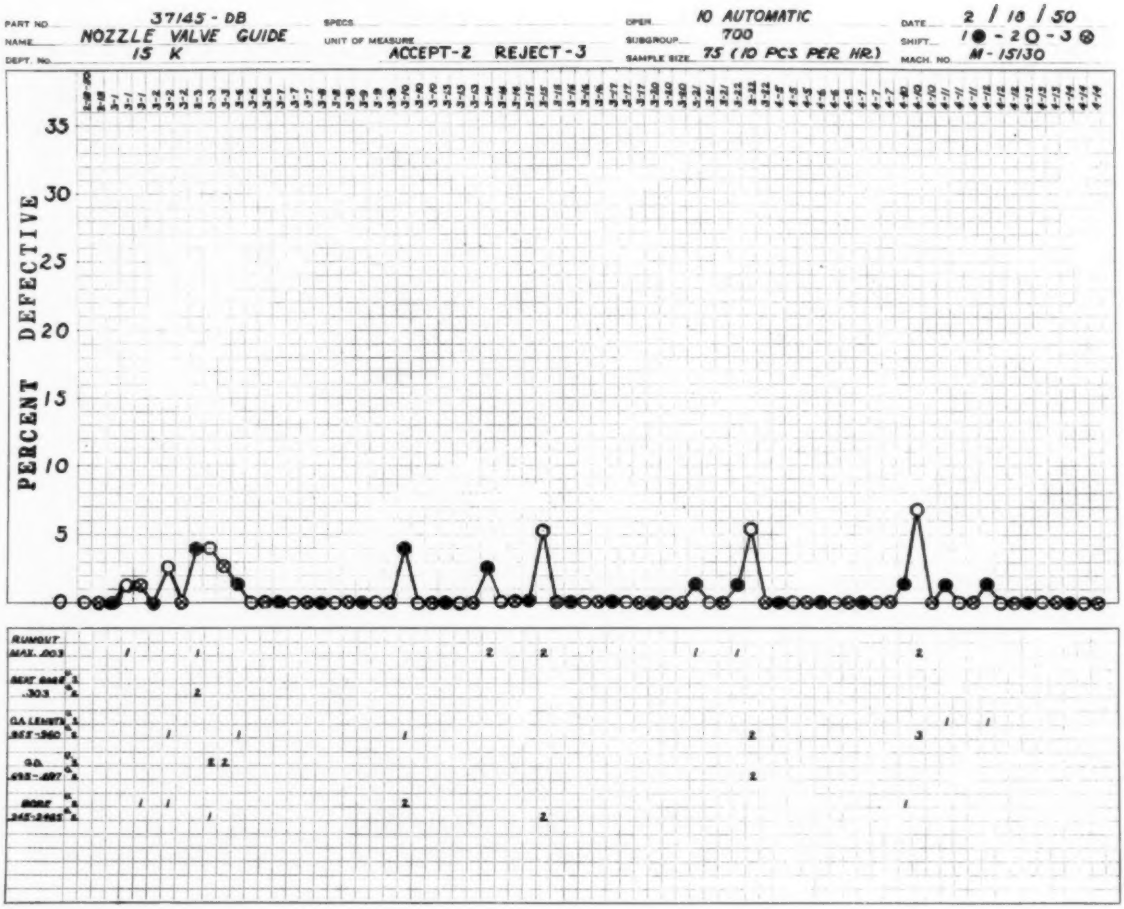


TAKEN FROM: NAVY DEPARTMENT STANDARD SAMPLING INSPECTION TABLES—APPENDIX X

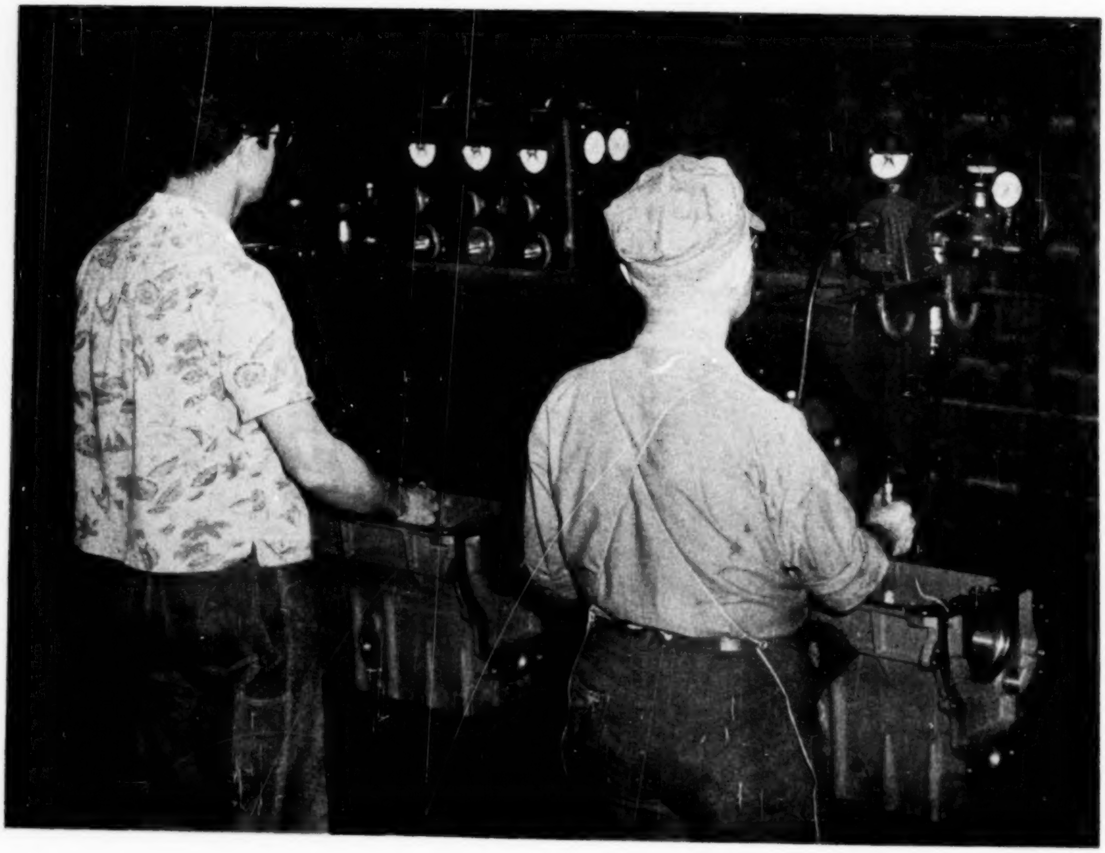
SLIDE NUMBER 2



SLIDE NUMBER 3



SLIDE NUMBER 4



Inspecting Silver Diamond Crank-cases

Indianapolis Works

SLIDE NUMBER 5

QUALITY RATING

KEY SHEET

MAJOR CHARACTERISTICS

RATING NO.	PERCENT DEFECTIVE
1	.00 - .50
2	.51 - 1.00
3	1.01 - 1.50
4	1.51 - 2.00
5	2.01 - 2.50
6	2.51 - 3.00
7	3.01 - 3.50
8	3.51 - 4.00
10	4.01 AND ABOVE

MINOR CHARACTERISTICS

RATING NO.	PERCENT DEFECTIVE
1	.00 - 1.50
2	1.51 - 3.00
3	3.01 - 4.50
4	4.51 - 6.00
5	6.01 - 7.50
6	7.51 - 9.00
7	9.01 - 10.50
8	10.51 - 12.00
10	12.01 AND ABOVE

NO. 9
Price 25¢

INSTALLING A COMPLETE QUALITY CONTROL SYSTEM

by

Stanley C. Amren
Director of Quality Control
The United States Time Corp.
Waterbury, Connecticut

for presentation at the

FOURTH NATIONAL CONVENTION
and
FIFTH MIDWEST CONFERENCE

of the

AMERICAN SOCIETY for Quality Control

June 1 and 2, 1950
Milwaukee Auditorium, Milwaukee, Wisconsin

FIFTH MIDWEST QUALITY CONTROL CONFERENCE
P. O. Box 1204, Milwaukee 1, Wisconsin

INSTALLING A COMPLETE QUALITY CONTROL SYSTEM

by

S. C. Amren - Director of Quality Control

U. S. TIME CORP.

Fellow - Southern Connecticut Section - ASQC

The task of installing a complete Statistical Quality Control Program in any sizable industry presents many problems and considerations. Top management naturally expects returns on it's investments and S. Q. C. should be considered as such. From this investment is expected better operating efficiency, lower scrap and rework losses, reduced inspection costs, and better customer relations. Unfortunately, S. Q. C. is not something that can be purchased in a package ready for use, but requires considerable perseverance for success. Some industries have failed to realize a profit from it's investment in S. Q. C. and invariably this failure can be traced to a lack of wholehearted cooperation on the part of management.

The problem from the outset is for the Quality Manager to sell management thoroughly enough so that there is some enthusiasm behind the movement all the way down through the lines of supervision. As soon as this has been accomplished, he must evaluate the overall situation and attempt to direct his efforts toward a successful goal. Probably his initial considerations should be given to at least all of the following.

1. Does he have the necessary technical personnel to administer the program?
2. What is the present cost of inspection?
3. What is the calibre of the present inspection force?
4. What is the present cost of scrap and rework?

5. What is the cost of servicing and replacement of field returns?

6. Does the inspection department have adequate measuring equipment and gaging?

7. How should the training program be undertaken?

Consider the above items in some detail.

1. Does he have the necessary technical personnel to administer the program?

The success of a Quality Control program providing it is given the proper backing depends largely on the individuals who are to administer it. These individuals should be thoroughly trained in S. Q. C. techniques and should have either an engineering or statistical background, preferably both. If the industry is large enough to warrant several such men it would be well to obtain a mixture of Engineers and Statisticians as both such backgrounds are seldom found in one man. It is also helpful if these men are familiar with the products and processes involved and possess personalities which are conducive to cooperation.

2. What is the present cost of inspection?

It is important that cost of inspection be determined accurately using some reference. Probably the most common references are percent of direct labor and inspection cost per unit. It is advisable to obtain both as sometimes S. Q. C. reduces direct labor costs per unit more than it reduces inspection costs, and consequently the inspection cost shown as percent of direct labor will be unfavorable. If, however, you have figures showing inspection cost per unit you may avoid undue criticism. This information should be charted every week or month as the program progresses.

3. What is the calibre of the present inspection force?

It is not uncommon to find Inspection departments filled with older people who can no longer keep pace with mass production lines. If this is

9

the case, the program is doomed unless corrective action is taken. S. Q. C. changes the entire concept of inspection, and it is much easier to change it in a young progressive mind than it is in someone who has become accustomed to old fashioned methods over a long period of years. It is essential that inspectors have the ability to understand at least the basic principles of Quality Control methods, since lack of this ability will only cause them to disregard the sampling plans, and consequently no benefits can result.

4. What is the present cost of scrap and rework?

Another way of showing progress of an S. Q. C. program is by charting scrap and rework. It is important that accurate figures be obtained to show these losses before the program gets under way, since without this picture it may be difficult to substantiate the investment that management is making. It is not sufficient, however, to obtain scrap figures alone; it is also important to get production figures so that scrap and rework may be shown as a percent of good work.

5. What is the cost of servicing and replacement of field returns?

No product however good is free from defects and most industries carry a Customer Service Department to handle complaints and repair or replacement problems. The cost of maintaining this department plus replacement and servicing costs are items that Quality Control will reduce. It is advisable to work closely with this group as much valuable information is often found in it's files.

6. Does the inspection department have adequate measuring equipment and gaging?

Invariably, an S. Q. C. installation uncovers some instances of inadequate gaging and measuring equipment. This is one of the major contributions that these techniques make to inspection departments. Quarrels and misunderstandings brought about by questionable gaging are no longer necessary.

9

There is no way of positively assuring ourselves that a product meets specifications unless the measuring equipment being used has been statistically proved to be adequate. On the other hand, measuring equipment has sometimes been condemned when it has not been at fault. Such questions will be answered by statistical techniques.

7. How should the training program be undertaken?

The problem of training the people in various capacities in the organization is a large but necessary task. Probably the first group to receive this training should be the key inspection personnel. However, it soom becomes apparent that production supervisors and methods engineers need the training almost simultaneously with inspection personnel. Then as the program progresses into the machine capability stage and revision of some tolerances are indicated, it will be necessary to enlighten design engineers about the techniques being used. It is also advisable to acquaint your sales force with the new program so that they may take advantage of the more reliable quality aspect in talking with customers.

The foregoing has dealt with some of the earlier aspects that must be considered in an S. Q. C. program. However, this is only the beginning and the most difficult hurdle is with the production employees themselves. A great deal of effort should now be expended in selling quality workmanship to these people. This can best be done by the Quality Control Engineers, who are in constant touch with production. There will be occasions where an S. Q. C. application will show some startling results. The employees concerned with this improvement should be given more than their share of credit as instances of this nature will do more to promote the program than anything else. Be sincere with the employees, sell the program to the union, make everyone concerned feel the importance of quality workmanship for his own

9

security. Remind them often that quality makes sales, sales make jobs. Do not hide scrap figures but show them how much is being lost for this reason.

Most employees like to do good work providing someone at the supervisory level is showing an interest in what is going on. Quality Control charts afford a real opportunity for the supervisor to keep posted on what is going on with a minimum amount of effort. In cases that are out of control due to operator inefficiency make sure the operator knows what he is supposed to do and that he has the means of doing it properly.

EFFECT OF S. Q. C. ON VARIOUS INDUSTRIAL FUNCTIONS SUCH AS METHODS, DESIGN ENGINEERING, PRODUCTION, SALES, PURCHASING AND INSPECTION.

METHODS

The major contribution that S. Q. C. makes to the methods department is in showing process capabilities. This is a great aid to the Methods Engineer since it is now possible to accurately establish the fitness of any machine or process for its intended job.

DESIGN ENGINEERING

Design Engineering embodies the function of setting tolerances and establishing specifications. Often in the past these have not been realistic, causing undue hardship on both methods and production. With S. Q. C. the design engineer now has available to him a wealth of information on what can be expected of men, machines, and material. He is thus in a better position to establish realistic specifications for the best efficiency and quality balance.

PRODUCTION

This function probably bears the brunt of the attack by S. Q. C. However, it also realizes more benefits than any other function providing production supervisors take advantage of the information being made available.

9

Weaknesses of almost any nature are pointed out and corrective action results in substantial savings in improved efficiency, scrap, rework, and better quality.

SALES

The sales organization is always looking for high quality to keep customers happy. S. Q. C. is without a doubt the best ally that any salesman ever had since he is assured that the Quality angle is being controlled scientifically and with much greater dependability. He is proud of this fact and uses it to good advantage on his customers. Furthermore, an increasing number of purchasing agents are asking vendors if their product or material is statistically controlled before buying.

PURCHASING

One of the functions of the purchasing department is that of supplying the industry with raw material, parts, and various other items at the most advantageous prices. It is sometimes difficult for the purchasing agent to know whether or not he is getting a bargain. S. Q. C. affords accurate comparisons of vendors as well as more accurate information about the material being purchased.

INSPECTION

The function of an inspection department is to assure management that the product being made conforms to specifications and acceptable standards of workmanship. This has always been the function of the inspection department, but not until the advent of S. Q. C. has it been possible to adhere to it. Inspection might better have been described as the department that did all the sorting, thus covering up the sins of the production departments. Under modern techniques, inspection does no sorting whatsoever. This must be done or at least be charged to the department making it necessary. Sorting is no more a function of inspection than rework is, since both are making

9

the material fit for use. Why then shouldn't it be paid for in the same manner.

This also holds true for purchased material. If purchased material is unsatisfactory there are three alternatives.

1. Reject the material.

2. Sort and reject the defectives.

3. Accept the material with production taking a variance.

The first of these is the ideal, but unfortunately due to urgent situations it can not always be done. If it is necessary to resort to 2 or 3, the sorting costs or variance incurred should be charged to purchasing. Purchasing may or may not be successful in recovering part or all of the costs from the vendor and in the event that they are not recoverable they should be charged to the original purchase cost. In this way the true cost will be apparent and it may be more economical to buy a better grade of material.

THE PROGRAM AT U. S. TIME CORP.

It is now 15 months since we began our S. Q. C. program at U. S. Time and we have followed the thinking outlined in the foregoing quite closely. As Director of Quality Control I report directly to the president of the company and am responsible for all S. Q. C. and Inspection activities in all four of our plants. We have been given complete backing, from top management, in our efforts with the result that our progress has been quite rapid. We have gained considerably in production efficiency, better quality, reduced scrap and rework, etc. but it is still a little early to determine what the effect will be on our customer service costs. We believe, however, that a large portion of the present cost of the entire quality function will eventually be paid for through the reduction of these costs.

9

This company manufactures a variety of inexpensive watches, alarm clocks and various other timing devices. Because of the production volume and specification requirements on watches we have directed most of our attention to this item.

As our training program got under way it soon became obvious that the educational background of most of our Inspectors was such that statistical techniques could never have been absorbed. This was a major obstacle but some sort of action was necessary, and consequently the problem was discussed with our Director of Industrial Relations. It was finally decided to discuss the matter with our union and make a proposal that we would give an aptitude test to all employees who had Inspection experience. The test was to be in two parts; one for straight arithmetical ability and the other for intuitive ability. The union agreed to the proposal and notices were then sent to all who had inspection experience including those who had been laid off due to production cutbacks. Approximately 250 candidates turned out and out of this group 55 qualified. These were immediately put to work and those who failed, but had enough seniority were transferred to various jobs in the production and repair departments. We have since found it necessary to hire additional inspectors, but have insisted that these men pass this same test and have had inspection experience. As a result of these steps we now have a high grade progressively minded inspection department, and this has contributed materially to our progress.

An additional investment was required for measuring equipment since some of these items were found wholly inadequate. We now have some of the finest in the world and even the best is none too good considering the tolerances we must cope with. Our gaging is continually being improved to fit S. Q. C. requirements.

Early in our program we were searching for a difficult and trouble-

9

some process to which we might direct our efforts and gain some prestige with production. The production manager was happy to point out a screw machine job requiring 64 hours of machine time per day plus the time of 6 girls to sort the parts, whereas 16 hours of machine time should be sufficient providing a 100% yield were obtained. We immediately began a series of machine capability tests using $\bar{\bar{X}}$ and R charts on the dimensions causing the difficulty. It was soon obvious that the machines being used could not yield 100% with present tolerances. We then had our first encounter with Engineering in regard to tolerances. We explained to them that by increasing the tolerance in such a direction that on the average we would have a closer fit between the part in question and the mating part. Engineering was quite concerned that some of the parts would not go together since this permitted a line to line fit. After considerable explanation of the laws of probability, the change was granted with the result that we now have only 16 machine hours per day and no sorting whatever. Furthermore, the assembly department is pleased with the better average fit. This application did give us a tremendous boost with the result that several more jobs of similar nature were called to our attention. Machine capability tests are continually being run on new jobs so that discrepancies like the above can be ironed out before any damage can be done. No new job is permitted to run on a production basis until our capability tests indicate that it can be made to specifications.

Out inspection plans in the various departments are tailored to fit the requirements of each group of machines. In departments, such as Automatic Screw Machine, Pinion Cutting, and Secondary Operations, we employ three different plans.

1. Acceptance sampling of the entire days production.

This type of plan is the most economical and is used where very

little difficulty is encountered in meeting print requirements or where operators are extremely capable and scrap is seldom produced. The sample size of course depends upon the lot size and the particular requirements of the part.

2. Process Inspection based on attributes.

   This type of plan is employed where more constant supervision is required and scrap could readily be produced. The inspector makes a complete round every 2 hours and takes a first sample of 10 with c = 0. If c = 2 or more, the production during that two hour period is scrapped. If c = 1, a second sample of 20 is taken and if any additional defects are found, the lot is scrapped. If no more defects are found the lot is accepted. This plan may seem lenient mathematically, but we feel it is a better balance between quality and production than our first plan, wherein the sample sizes were 20 and 40 and using the same "c" values. Furthermore, our present system permits the inspector almost twice as many rounds per day which is also desirable.

3. Variables sampling.

   There are always certain dimensions of critical nature and where machine capability tests indicate that the machine is barely able to meet print requirements. It is necessary in such cases to resort to $\bar{\bar{X}}$ and R charts to assure that no scrap is being made and to indicate in advance when machine settings need be changed.

The amount of Inspection required in departments such as the above depends on the ability of the operators and the number of machines they are required to run. It is folly to load an operator down with so many machines that he does not have time to keep checking the work being produced. Poor

operators or operators who are loaded down with too many machines require more inspection, thus creating an incorrect and unjust picture of direct and indirect labor charges.

Our sub and final assembly departments require the same kinds of plans as those used in our machine departments but to a lesser extent. There are a few processes requiring $\bar{X}$ and R charts because of their critical nature, but on the average the cost of Inspection in this area is much less than in our parts manufacturing areas.

We have made serious efforts to keep our Inspection costs to a minimum and still maintain the degree of control necessary for the most economical balance. It is interesting to note that in order to accomplish this it has been necessary to supply an inspection force to our parts manufacturing departments which is equivalent to 30% of direct labor. However, the cost of inspection in our sub and final assembly departments is less than 10% of direct labor giving us an overall average of 15%. To some, this may seem high but it should be borne in mind that our parts are extremely small with very close tolerances and consequently, difficult to inspect. Furthermore, it is to be expected that these percentages will be reduced as our program progresses and operators become better acquainted with these new methods. The ultimate will only be reached when operators become capable enough so that acceptance sampling can be used almost exclusively.

Since we purchase a variety of parts from outside sources it is also necessary to assure ourselves that these components satisfy our print requirements as well as those made in the plant. All such parts are subjected to an acceptance sampling plan which varies with the requirement of each particular part but rarely is in excess of 2%. Ordinarily, the parts are rejected if they exceed our established percent defective but we do have situations where due to emergencies it is necessary to sort them.

9

In cases of this nature we are careful in seeing to it that costs thus incurred are charged to purchasing. If these costs can not be recovered from the vendor, then it is only logical that they should be added to the original cost. In this way purchasing is in a better position to evaluate the various suppliers from the cost versus quality standpoint.

Our Sales organization has been very interested in the progress of our program and we have attempted to give them as much information as possible on the improvement in quality. Furthermore, we have run tests on competitive products and evaluated the results to determine our own weaknesses as well as our competitors strong points. Recent comparisons indicate that we now have the highest quality timepieces within our price range and we intend to keep ourselves in this position by a progressive program searching for continued improvement.

The results we have obtained thus far are substantial and quite gratifying. The percent defective from our screw machine department has been reduced from more than 5% to less than 2%. The production on one of our watch movement lines has been increased by 76% during the past year. This improvement was partially due to an incentive plan but the overall improvement could not have been realized without the improvement in parts the line is now getting. The percentage of watches from the movement assembly line accepted for the first 24 hour run has increased from 76% to 93.3%. The percentage lost during the 24 hour run has been reduced 28% and the total lost in the dialing and casing section has been reduced 45%. More important, however, is the improvement in quality of product as it leaves the plant. According to our Quality Analysis, the percentage of watches containing critical or major defects has been reduced from approximately 5% to less than 1%. We believe that this quality level is the best that has ever been achieved by any manufacturer of inexpensive non jewelled watches.

9

This quality level should reduce our customer service and repair expenses considerably and help to defray the cost of our program to a great extent.

SUMMARY

When installing a Quality Control Program the first requirement is a complete picture of the present status in regard to scrap, rework, inspection cost, customer repair service, and manufacturing efficiency. Then a thorough job of selling the program to top management and all department heads that will be affected is the next requirement. In selling the program it is necessary to do some educating so that the individuals concerned have at least an understanding of the basic fundamentals involved. Be sure that the inspection department is mentally capable as these techniques require some arithmetical ability. Select and train qualified men as Quality Control Engineers. Proceed with the program cautiously tackling the most troublesome situations first. Take advantage of spectacular cases to further the cause but be sure to give credit to production workers when due. See to it that corrective action is forthcoming where control charts indicate the necessity.

Make every effort to show methods and design engineering the power of these new techniques as these function more than anyone else will appreciate the factual information made available through process and machine capability tests.

Keep up to date charts showing the progress that is being made with regard to scrap, rework, inspection cost, customer repair service, manufacturing costs and outgoing quality levels. These should be reviewed periodically with top management as well as all supervisors concerned.

9

When the program is well under way and results substantial it is advisable to prepare a report for top management showing in dollars and cents what savings have been affected.

9

NO. 10
Price 25¢

STATISTICAL QUALITY CONTROL OF CLERICAL AND MANUAL OPERATIONS

by

James M. Ballowe
Manager of Inspection
Alden's, Inc.
Chicago, Illinois

for presentation at the

FOURTH NATIONAL CONVENTION
and
FIFTH MIDWEST CONFERENCE

of the

AMERICAN SOCIETY for
Quality Control

June 1 and 2, 1950
Milwaukee Auditorium, Milwaukee, Wisconsin

FIFTH MIDWEST QUALITY CONTROL CONFERENCE
P. O. Box 1204, Milwaukee 1, Wisconsin

RESULTS OBTAINED DURING FIVE YEARS
OF OPERATION IN A MAIL ORDER PLANT

James M. Ballowe*

Statistical quality control can be applied to any kind of a process in which there is repetition. Because of this, a mail order plant is a statistical mecca. With many thousands of orders each day, and with the processes broken down into activities so that each employee does just a certain amount of work on the order and no more, there is a situation which lends itself to statistical quality control.

Most quality control applications pertain to machine operations. While statistical formulas and quality control charts are available, not many applications have been reported on purely clerical and manual work. As distributors, we do no manufacturing. Clerical and manual operations are the only kind we have. In introducing the program, therefore, we were faced with some of the practical problems which always arise in the adaption of a new management technique.

Since statistical quality control had been used so successfully on machine operations, my original thought in the fall of 1944 when I was taking the first 10-day intensive course in the subject at the State University of Iowa, was to draw an analogy between machine operations and human operations. While one dislikes to compare humans with machines, we might recognize that some similar characteristics exist as shown by the following analogy:

10

Machine Operations	Human Operations
Scrap (or rework)	Throwbacks (for correction)
Repetition of defective parts	Repetition of same type of error
Specifications	Goals
Major and minor defects	Errors affecting and errors not affecting the customer

Introductory Problems

For any management technique to be applied successfully, there are usually practical operating problems which have to be solved. I knew the following hurdles would have to be overcome if statistical quality control methods were going to be applied successfully in our plant.

1. The simplest type of mathematics. While it is impossible to imagine statistics without any mathematics, I knew that the department managers and supervisors who would operate these new control methods were not mathematicians. Nor do they have the time to work out certain necessary techniques such as control limits. To overcome this objection, we prepared a quick-reference table as shown on page 3. After a manager has observed his error ratio, he merely has to look at this table to determine his control limits.

2. The control chart would have to be simple, easy to understand, and quick to comprehend. We prefer to do most things "at a glance".

3. The charts would have to be practical from an upkeep point of view but attractive in appearance. Nothing creates a better atmosphere in an operating department than charts which are kept up, and which are neat and attractive in appearance.

All of the above problems were solved in making our application of statistical quality control. After I describe our procedures and explain how we maintain our control charts, you will readily see how we were able to solve the practical problems I have enumerated.

VALUE OF THE CONTROL LIMITS OF AN np-CHART CORRESPONDING TO VARIOUS VALUES OF AVERAGE NUMBER OF ERRORS MADE PER 100 WORK UNITS

CENTER LINE (Average Number of Errors per 100)	UPPER CONTROL LIMIT (2Sigma Limit)	CENTER LINE (Average Number of Errors per 100)	UPPER CONTROL LIMIT (2Sigma Limit)
.10	.73	2.10	4.97
.20	1.09	2.20	5.13
.30	1.39	2.30	5.30
.40	1.66	2.40	5.46
.50	1.91	2.50	5.62
.60	2.14	2.60	5.78
.70	2.37	2.70	5.94
.80	2.58	2.80	6.10
.90	2.79	2.90	6.26
1.00	2.99	3.00	6.41
1.10	3.19	3.10	6.57
1.20	3.38	3.20	6.72
1.30	3.57	3.30	6.87
1.40	3.75	3.40	7.02
1.50	3.93	3.50	7.18
1.60	4.11	3.60	7.33
1.70	4.29	3.70	7.48
1.80	4.46	3.80	7.62
1.90	4.63	3.90	7.77
2.00	4.80	4.00	7.92

THE LOWER CONTROL LIMIS is zero for each of these cases.

Aldens Quality Control

10

The Alden Control Chart

The np chart, used at Aldens, measures 15 inches in height and 10 feet in length. It is prominently displayed in the center of the activity being controlled as shown in the photograph on the next page. On the bottom of the chart we show the days of the month with sufficient space for each day to allow for about six plottings. In the left margin of the chart the number of errors is indicated, from 0 to 13. Parallel lines running horizontally across the entire chart aid in observing the number of errors found in the sample.

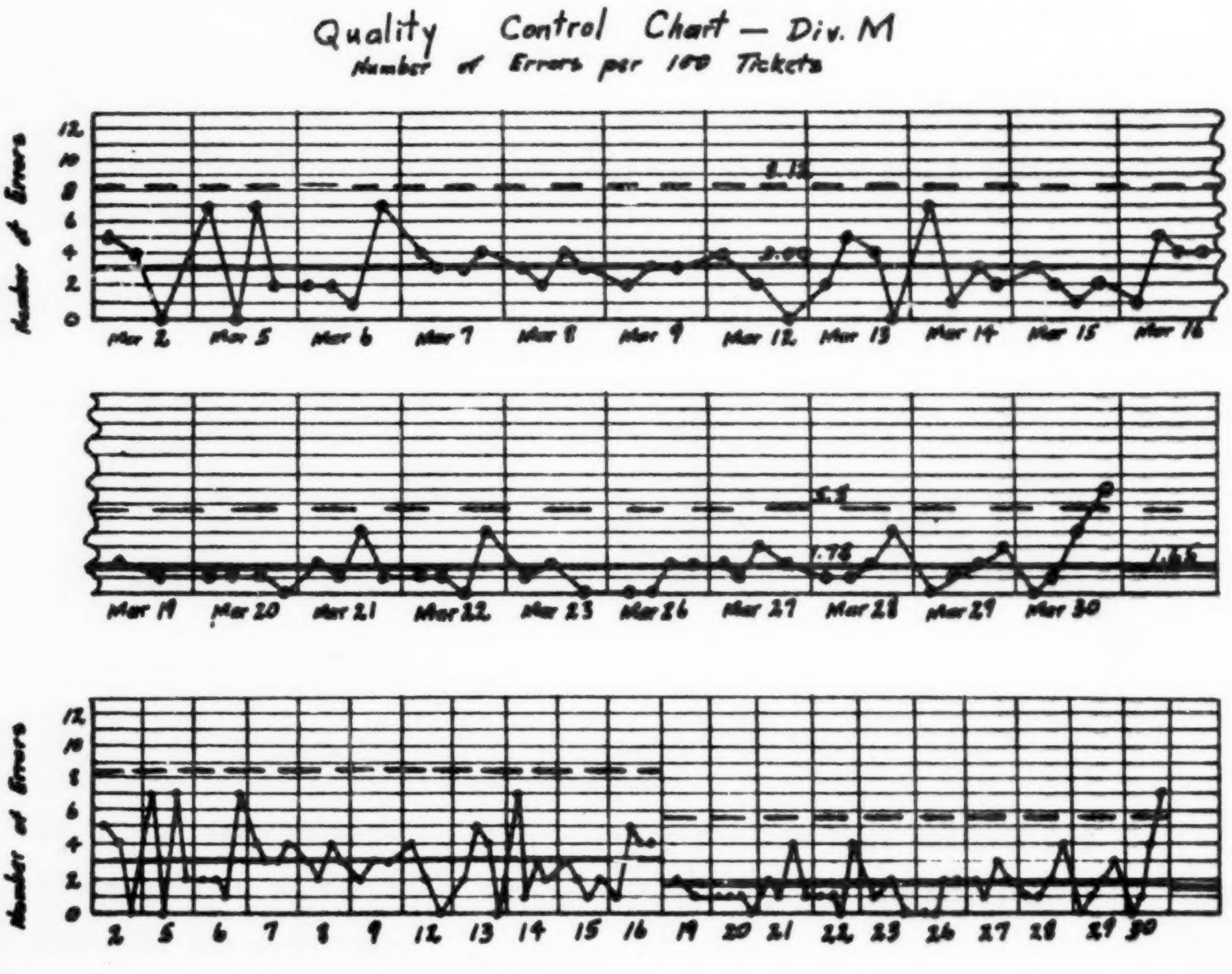


Figure 1

The sample size is 100 work units in all activities. In some departments the work unit is a customer's order; in the merchandise departments it is called a "ticket"; in the Shipping Room the work unit is an entire package. The advantage of using a sample size of 100 work units is that the error ratio contained in the sample is easily computed and readily comprehended.

In making the plottings on the chart we use circular gummed tabs one-half inch in diameter. These are affixed to the chart on the horizontal line which indicates the number of errors found in the sample lot inspected. We then superimpose on the tab a gummed numeral indicating the number of errors the line represents. These tabs are connected by means of red gummed paper tape 3/16 of an inch in width, easily affixed in the same manner. This construction is indicated below:

Figure 2

By using these materials, all of the charts are kept uniform and attractive in appearance. Anyone who can apply a postage stamp can make the plottings.

10

We maintain the charts on a monthly basis. On the first day of each month we put up a fresh chart in each department, using a staple gun. The sheet is affixed to a permanent backboard. In the upper right corner of the chart we show the error ratio established for the preceding month. In the upper left corner we show the upper control limit applicable to that particular activity. So long as the plottings remain below the upper control limit, with reasonable variations, we have confidence that the process is consistent. When a plotting falls outside the upper control limit, however, a directive from top management requires that the department manager personally inquire into the result to see whether there is an assignable cause--whether a particular operator or type of error is responsible. If the plottings stay within the control limits, we do not require that the manager give attention beyond his normal review of production results.

A solid line, extending across the chart, shows what the goal is for the month. In setting a goal we consider recent experience, the possibility of major volume fluctuations, and the results being obtained on comparable operations in other departments. We advise against setting a goal when a chart is first placed on a new activity. At that stage the important thing is to find out how the process is really running. We have gone as long as two months before setting a goal. But after sufficient data is obtained, a reasonable goal, possible of attainment, should be set.

In the mail order business there are wide volume fluctuations, particularly prior to Christmas and Easter. Therefore, it is

advisable to set a goal which will make a reasonable allowance for variations. The goal may be set slightly higher or slightly lower. The influx of a large number of people or the change of a technique would require a review of the goal established.

The Economical Level

It would be amiss for me to imply that we expect perfect work. We know our employees are going to make errors. We also know it would be far too expensive to try to eliminate all of their errors. In fact, when an activity reports a substantial reduction in errors and we are convinced the results are honest, our first consideration is how much we can reduce inspection. It remains for top management to determine the economical level, after a consideration of cost and possible customer reaction.

But what is more important and what we are anxious to pay for is the assurance of the constant protection which our quality control program affords. We want to know immediately when we should focus attention on results which are not normal to a process.

A further description of our quality control procedure may be desirable before going into actual applications. As we all know, a quality control chart in and of itself will not reduce errors. The remediable action taken after data is obtained will be primarily responsible for any improvements.

To get these data in good form at Aldens we use what we call a data sheet and an analysis sheet. A separate data sheet is used

10

for every 100 units of work inspected as shown on page 10. On this sheet we show a breakdown by operator and by type of error. The 100 work units are a blend of a proportionate amount of work from each employee in the activity. For example, if there are 10 employees we select approximately 10 work units from each. The number of work units inspected on each employee is shown on the top half of the data sheet. On the bottom half we describe each error found and show the operator responsible. Also, the individual who made the mistake signs the sheet to indicate she was shown the error. We are great believers that people like to see the mistakes which they make. This is much better than being told, "You're making too many errors...cut 'em out!"

The results shown on the individual data sheets are then transposed to a weekly analysis sheet which shows the total number of work units inspected on each operator for the week, together with a breakdown of the types of errors made, as shown on page 11. We have found that what appears to be an isolated type of error on any one day generally falls into a definite error category or pattern when all errors are accumulated over an entire week.

The weekly analysis sheet is sent to the Operating Superintendent for review. A copy is retained in the activity for study by the trainer and supervisors. So much for the mechanics of our program; now to applications.

First Application

Our first quality control application was made in the spring of 1945 in one of the merchandise departments on the order-picking and checking operation. Order pickers walk or skate up and

10

Sheet No.________

ALDENS QUALITY CONTROL

Data Sheet

Department________________ Date________________ Inspector________________

Operator No.	Units of Work Inspected

Description of Errors

Error No.		Operator Responsible

When you complete the inspection of 100 work units, post the number of errors found on your control chart **immediately.** If an error is serious or repetitive in character, call the matter to the attention of your supervisor **at once.** This worksheet to be circulated to and initialed by the following:

Department Manager ____________ Supervisor____________ Trainer ____________

F1053

Aldens QUALITY CONTROL
ANALYSIS SHEET

Section ______________________ Week Ending ______________________

	Operator												
	Date												
UNITS OF WORK													
	Total												
	Operator												
TYPES OF ERRORS	Catalog No.												
	Color												
	Size												
	Quantity												
	Price												
	Stamp												
	Total												
	% Errors												

Entries should be made on this recap sheet daily, with data taken from your Quality Control work sheets of the day previous. This report to be made in duplicate and the original sent to your Department Manager each Monday, for the week previous.

____
Section Manager

10

F1207

down stock aisles selecting merchandise by catalog number, size, color and quantity. These work units are then taken to a conveyor line where another employee checks to see whether the selections are correct, after which the merchandise is placed on a conveyor belt which carries it to a spiral gravity chute dropping to the shipping room.

The point of inspection is at the end of the conveyor belt at the entrance to the gravity chute. This location is a natural place to afford a random inspection. The errors found indicate the quality of work being performed by both the order picker and checker.

Upon completing the examination of 100 work units, the inspector goes to the control chart and posts the number of errors found in the sample. This result must be posted immediately, otherwise much of the value of the program is lost.

Prior to the use of statistical quality control in this one merchandise department, we predetermined that a 3.00 per cent error ratio existed. The process was within statistical control but at an unsatisfactory level, as shown in the first half of Figure 1. As a goal, we set an average error ratio of 1.78 per cent, asking for a 40 per cent improvement. Actually, the error ratio dropped within two weeks to 1.65 per cent, a 45 per cent improvement as shown in the second half of Figure 1. During the following months there was a further reduction to .70 per cent - - a quality level which was maintained during subsequent months. This improvement from 3.00 per cent to .70 per cent was attributed to the enthusiastic support given to the new quality control program by the department manager, his supervisors

10

and employees. Special efforts were made to find errors, study errors, and to eliminate them at their source. The improvements made were verified by an independent check of outgoing packages. Adjustments with customers decreased accordingly.

It is also interesting to observe that the production efficiency of this department increased from about 82 to 107 during the period the quality performance improved. We believe that when errors are once brought under control and the results show a good performance is being maintained, there is a natural tendency to speed up production.

A question often asked is why quality control should increase production. For the answer to this question we turn to member of the fair sex. While there are not many young ladies working on quality control programs, there's one who hears a lot about mine and she can provide us with the answer to the question.

For several months after she learned to pull herself up to a standing position, my one year old daughter wouldn't let go. She was lacking in confidence. Once she acquired that confidence, away she went.

Confidence in anything we do encourages us to do more. Whether it's riding a bicycle, learning how to typewrite, or if it's operating a production line in a plant. When we have the assurance that quality is being maintained at a satisfactory level, there is no reason to withhold production if greater speed is possible.

After the success of this first installation, other merchandise department managers requested a similar program. It was

10

gradually expanded into all of the merchandise departments, the billing, shipping, exchange and returned goods departments.

During the four year period beginning in January, 1946, and ending in December, 1949, the combined error ratios for all the merchandise departments was reduced by 57.9 per cent. It is believed that statistical quality control played a significant part in bringing about this improvement. A few of the departments now have an error ratio as low as one tenth of 1 per cent, although our over-all plant objective is now one-half of 1 per cent.

General Office Applications

In the fall of 1945 the General Office Manager requested that statistical quality control be introduced in all of his departments. Clerical work performed in our General Offices includes the following:

1. Open envelope, remove contents, verify remittance, apply cash impression to order blank.

   Error possibilities: Total remittance incorrectly; apply wrong cash impression; fail to transpose customer's name and address from postal money order or personal check when information is missing from order blank and envelope.

2. Read order to see whether any phase of transaction will not be handled in regular mail order process. If so, apply special rubber stamps, make abstracts on special requests, inquiries and complaints.

   Error possibilities: Fail to handle special phase of transaction.

3. Record order on customer's index stencil; show date received and amount. Imprint stencil showing customer's name and address on shipping label and order.

4. Pull entry "ticket" for each catalog number ordered; circle size, color and quantity wanted.

10

Error possibilities: Pull wrong tickets; circle wrong color, size or quantity.

5. Schedule order for shipment. A machine operation showing time, aisle, section, and bin number in Shipping room where order will be packed.

   Error possibilities: Switch tickets and send order to wrong customer.

All of the above activities are under statistical quality control. Applications have also been made on key-punch operations in the General Offices and on freight and express routing activities. During the four complete years statistical quality control has been in use in the General offices, 1946-1949, the combined error ratios have been reduced by 67.6 per cent.

In one clerical department, the Order Reading, where we originally had only one chart, we found we needed a still further breakdown by order classification, such as cash orders, no-cash orders, and credit orders. Accordingly, we now have charts on all of these separate activities. Also, in the General Office departments where we have both day and night shifts, we have a separate chart for each shift. As is generally expected, the night shifts started out with higher error ratios than the day shifts, but as the night force became more stabilized we now have better work in some activities than on the day side. When each shift knows what the other is doing, competition becomes possible. The chart on the wall for each shift speaks for itself.

About a year and a half ago statistical quality control was introduced into the Credit Department, where charts are now maintained on posting-checking operations, credit approval, follow-up typing,

10

files and related activities. Comparing January, February and March, 1950, with the same three months last year, combined error ratios for all credit department activities have been reduced 37 per cent.

The manner in which data is obtained for the control chart on the filing operation may be interesting. Before signout of work is made to file clerks duplicate stencil impressions showing customer's name and address are removed from the papers. After 100 names have been accumulated and the papers filed, a lookup is made to see whether they were filed correctly.

Monthly Quality Control Meetings

Each month we hold a quality control meeting in which all of the charts for the previous month are reviewed. This meeting lasts about one hour and is attended by all department managers, the plant superintendent, and the Director of Mail Order operations. All of the charts are reviewed and department managers are called upon to discuss any unusual results, whether unusually good or unusually bad. Supplemental efforts to go "beyond the chart" are described by managers who have found new devices helpful.

A definite amount of pride and assurance results when a manager with good results knows his charts are going to be seen by top management at this monthly meeting. Also, at this time, each manager has the opportunity to see all the other charts and hear discussion of them.

This raises the question of the assignment of responsibility for quality control. In our plant we feel that the control of quality

10

is just as much a managerial function and responsibility as the control of production, unit cost, absenteeism, or any other administrative control. For this reason, each of our department managers acts as his own quality control manager. He is personally responsible for the quality of work his people turn out. To take away this responsibility and transfer it to the Training Department or to the Inspection Department would be to relieve the manager of one of his most important functions as a manager, namely, to turn out quality production.

Beyond The Chart

The question of how far the manager should go to accomplish a necessary improvement is left to his discretion. Presumably, he knows his people best, and takes the action he deems necessary under the circumstances. Such questions as to the publicity which should be given to individual error makers, the penalties which should be imposed for poor work, the special awards which should be given for good work, are left entirely to his judgment.

Some managers attach the individual data sheets to the bottom of the control chart. Other managers have individual error slips made out, signed by the employee responsible, then posted on the control chart. Another places a gold star besides an operator's name on a bulletin board for perfect work done during an entire week. Usually, a plant-wide program has to be given a "booster" shot from time to time and for this we depend upon the originality of the managers working in the program.

The department manager must be constantly searching for ways

10

to eliminate error hazards. We know of no better way to reduce errors than to eliminate them at their source. In clerical operations, particularly, we have found that more opportunities exist than one realizes. For example, an error hazard is always present when figures are transposed. In our Exchange Department the practice had been to carry forward a "total settlement" figure from the space on the back of the returned goods billing jacket to the flap portion on the front of the papers. The Refund Department would then issue a bank check to correspond with the amount entered on the flap. This type of error wasn't numerous but the hazard was always present and an error was made occasionally. Rather than risk an error in transition, the Refund Department now works directly from the "total settlement" figure, entirely eliminating the error hazard at its source.

In order filling operations we have found that proper lighting, good bin numbering, and improved bin-tag color combinations eliminate error hazards. To reiterate, when an error condition exists, attention should *first* be directed towards completely eliminating the error hazard. If this cannot be accomplished, then the *second* consideration should be how to reduce the error ratio.

Say It In English

When symbols or abbreviations are used in plant operations, a good practice is to make them self-explanatory. Symbols or signals may work all right with experienced employees, but they needlessly complicate the training of new people. To illustrate, for years "XXX" on our printed entry ticket meant "time payment order." The word "credit" or "T.P." do not need any interpretation. A pink slip inserted

10

vertically to extend below the jacket flap meant "Send customer a catalog." "P.P." meant "Postpaid." In training new employees it is much better to have expressions which say exactly what they mean.

$\bar{X}$ and R in the Measurement of Yard Goods

There is only one activity in our entire plant on which we use $\bar{X}$ and R. This is in the measurement of yard goods. Most of our piece goods are stocked in bolts and sold by the yard, the sales running into many thousands of yards daily. The customer may purchase any quantity desired but not less than 1/8 yard. Each order is cut to measure. Although we would like to cut the exact length, we find the natural tendency is to over-cut.

Prior to the introduction of statistical quality control on the cutting operation, an accumulation of samples indicated the average over-cut was 1.97 inches per cut, with both the average and range very erratic. The process was not in statistical control. We found that the order picker with the highest operating efficiency and the one thought to be the best of 20 such employees was giving away, occasionally, as much as 13½ yards of material an hour.

We found two principal causes for over-cuts...uneven tears and too many thumbs. By concentrating on both these factors we have been able to reduce the average over-cuts to 1.14 inches with further reductions indicated. Also, the process is now consistently within statistical control. A reduction of one inch per cut in the yard goods department is equivalent to a savings of between $15,000 and $20,000 per year.

10

Samples are taken in lots of 5 pieces from each operator, at random throughout the day. The samples are then measured for exact length. The average of the over-cuts found in the sample is plotted on an average, $\bar{X}$, chart; and the variation of the over-cuts found in the sample is plotted on a range, R, chart. Each plotting is identified with the name of the employee responsible.

A Five Year Evaluation

After five years of application in our plant, statistical quality control is now considered an integrated part of our standard operating procedures, as is time and motion study, plant lay-out, job evaluation, or any other management control. The appraisal of the value of the methods of our department managers is best illustrated by one whose quality of production is so high there is rarely an error plotting on his control chart. Yet, he refuses to discontinue the procedure. He states that "The chart is the best advertisement I have."

To summarize what we believe to be the outstanding advantages of the methods:

1. The approach is analytical. The randomness of the sample, the significance of variations in the sample and a point out of control, the classification of the data by employee and type of error, and the psychological effectiveness of the chart all add up to the best known scientific approach.

2. Variation is recognized. Whether or not we care to admit that variations will occur we know they exist. Employees and managers alike feel better about a production control when they know the results expected will be reasonable and uniformly applied.

3. Employee has confidence. The assurance that quality is being constantly maintained in the process builds confidence and permits increased production.

10

4. <u>The need for quality control is ever present</u>. The control chart is a reminder that the control of quality is an ever present problem, not one which can receive only occasional attention.

5. <u>Inspection results become productive</u>. Inspection results which formerly went into a desk drawer are now out in the open, on the control chart, where everyone can see what is taking place. A representative of management can tell at a glance:

   a. Whether inspection is being made daily.
   b. How much inspection is being made.
   c. The number of errors found in each 100 work units.
   d. The trend errors are taking.

6. <u>Employee reaction is favorable</u>. The employee knows what is expected of him and how he is doing while his work is being done.

A word about the sampling of incoming merchandise from supply sources. A manufacturer cannot expect his distributor to examine every piece of merchandise delivered to him, nor can the distributor afford to do this. How much better it would be if a quality control chart could accompany each shipment of merchandise, showing the quality which was maintained during the manufacturing process! This would reduce or entirely eliminate the need for sampling inspection by the distributor, a savings which could be passed on to the customer in the form of a lower price.

In my opinion, the day is not far off when purchase orders will read "Quality Control process chart must accompany merchandise." Manufacturers will proudly describe their products as "Produced under quality control." Everybody will know what everybody else is talking about and everyone will know what to expect.

---

*Manager of Inspection, Aldens, Inc.; Fellow, American Society for Quality Control, State University of Iowa Section.

10